Extreme Value Modelling with Application in Finance and Neonatal Research

A thesis submitted in partial fulfilment of the

requirements for the Degree of

Doctor of Philosophy

in Statistics

by Xin Zhao

Department of Mathematics and Statistics

University of Canterbury

2010

Abstract

Modelling the tails of distributions is important in many fields, such as environmental science, hydrology, insurance, engineering and finance, where the risk of unusually large or small events are of interest. This thesis applies extreme value models in neonatal and finance studies and develops novel extreme value modelling for financial applications, to overcome issues associated with the dependence induced by volatility clustering and threshold choice.

The instability of preterm infants stimulates the interests in estimating the underlying variability of the physiology measurements typically taken on neonatal intensive care patients. The stochastic volatility model (SVM), fitted using Bayesian inference and a particle filter to capture the on-line latent volatility of oxygen concentration, is used in estimating the variability of medical measurements of preterm infants to highlight instabilities resulting from their under-developed biological systems. Alternative volatility estimators are considered to evaluate the performance of the SVM estimates, the results of which suggest that the stochastic volatility model provides a good estimator of the variability of the oxygen concentration data and therefore may be used to estimate the instantaneous latent volatility for the physiological measurements of preterm infants. The classical extreme value distribution, generalized pareto distribution (GPD), with the peaks-over-threshold (POT) method to ameliorate the impact of dependence in the extremes to infer the extreme quantile of the SVM based variability estimates.

Financial returns typically show clusters of observations in the tails, often termed "volatility clustering" which creates challenges when applying extreme value models, since classical extreme value theory assume independence of underlying process. Explicit modelling on GARCH-type dependence behaviour of extremes is developed by implementing GARCH conditional variance structure via the extreme value model parameters. With the combination of GEV and GARCH models, both simulation and empirical results show that the combined model is better suited to explain the extreme quantiles. Another important benefit of the proposed model is that, as a one stage model, it is advantageous in making inferences and accounting for all uncertainties much easier than the traditional two stage approach for capturing this dependence.

To tackle the challenge threshold choice in extreme value modelling and the generally asymmetric distribution of financial data, a two tail GPD mixture model is proposed with Bayesian inference to capture both upper and lower tail behaviours simultaneously. The proposed two tail GPD mixture modelling approach can estimate both thresholds, along with other model parameters, and can therefore account for the uncertainty associated with the threshold choice in latter inferences. The two tail GPD mixture model provides

a very flexible model for capturing all forms of tail behaviour, potentially allowing for asymmetry in the distribution of two tails, and is demonstrated to be more applicable in financial applications than the one tail GPD mixture models previously proposed in the literature. A new Value-at-Risk (VaR) estimation method is then constructed by adopting the proposed mixture model and two-stage method: where volatility estimation using a latent volatility model (or realized volatility) followed by the two tail GPD mixture model applied to independent innovations to overcome the key issues of dependence, and to account for the uncertainty associated with threshold choice. The proposed method is applied in forecasting VaR for empirical return data during the current financial crisis period.

Acknowledgements

I would like to show my deepest gratitude to my supervisors Dr Carl John Scarrott, Dr Marco Reale and Professor Les Oxley for their guidance and support throughout my study.

A very special gratitude to Dr Dominic Lee for his stimulating feedback and advise on the Neonatal research and Bayesian inference. I would also like to thank Professor Michael McAleer, and Dr Felix Chan for their valuable suggestions at the MODSIM conference, and Dr Marcelo Medeiros for his inspirational discussion in realized volatility. A great honor to Doctor Glynn Russell in Christchurch hospital for his explanation of the medical background of Neonatal research.

I would also like to acknowledge Christchurch Women's Hospital in supplying the Neonatal data and Marcel Scharth and Marcelo Mederios for providing us with the high frequency intra-day data and a copy of their realized volatility data.

It is a pleasure to thank those who made this thesis possible. In particular, associate professor Jennifer Brown's spirit of encouragement in starting a PhD in statistics and the faculty members of the Department of Mathematics and Statistics at the University of Canterbury for all their support and help in my PhD studies.

Finally, I owe my special thanks to my husband Rong and my daughter Delicia for their constant support. The thesis would not be possible without their understanding and encouragement.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis primarily focuses on the modelling and forecasting rare events using "extreme value" based statistical models with applications in finance and neonatal research. Extreme value theory has been claimed to be capable of predicting unpredictable extremes, or even realizations outside the range of available data as Matthews (1996). Extreme value theory (EVT) has been applied in a broad range of areas with many applications in hydrology, pollution studies, material engineering, traffic management, economics and finance. Much of finance is concerned with measuring and managing financial risk or uncertainty in future returns, and such risk measurements include volatility as the standard deviation of the return and Value-at-Risk (VaR) as the maximum loss in a certain period. In this thesis, we examine the problems in applying extreme models in finance, such as threshold choice and dependence of extreme. New extreme value based models are constructed by integrating traditional extreme value models with other common econometric models, such as volatility model, to mitigate these issues in extrapolation and prediction of extreme returns. In the neonatal study, a stochastic volatility model with a particle filter is used in capturing the variability of the medical measurements of preterm infants. And extreme value model is applied in describing the tail behaviour of the estimated variances.

The motivation of the research is stated in Section 1.1. A brief review of the background and previous developments in extreme modelling in finance is outlined in Section 1.2. The objective of this thesis is given in Section 1.3. Section 1.4 describes the structure of this thesis, and the previous publications and papers relevant to this thesis are declared in Section 1.5.

1.1 Motivation

Extreme value modelling techniques have become widely used in the last 50 years in many disciplines, such as risk assessment of maximum loss in finance, extreme levels of a river in hydrology, and the largest claim in actuarial analysis (Coles 2001; Tancredi, Anderson, and O'Hagan 2006). The objective of extreme value modeling is to quantify the stochastic process of observations with small probability and even the extremes outside the scope of being observed. Modelling the tails of distributions is important in finance and the study of extreme events in risk management has long been important for investors and regulators. Extreme value models supply an asymptotic approximation for the tail distributions, which are very flexible in terms of the allowable tail shape behaviour. The attraction of the EVT based methods is that they can provide mathematically and statistically justifiable parametric models for the tail distribution, which can give reliable extrapolations beyond the range of the observed data.

Financial data is well known to exhibit relatively heavy (heavier than normal) tails, and extreme value theory has been shown to be a very useful tool in estimating and predicting the extremal behaviour of actuarial and financial products, such as predicting the largest claim in insurance and the Value-at-Risk (VaR), see Embrechts, Klüppelberg, and Mikosch (1997). However, applying extreme value models is not always straightforward and there are common issues in applications. The typical problem with extreme value models is the inherent sparsity of extremal data, which can result in the model identification and parameter estimation problem, particular with a complex model structure. The dependence of extremes is another common issue, particularly in financial applications. Under certain conditions, Beirlant, Goegebeur, Segers, and Teugels (2004) shows that a stationary process with short range dependence between the observations can also lead to the same extreme distribution family. However, the information contained in an extreme sequence with dependence is not as much as an independent identically distributed sequence. The statistical inference therefore needs to be adjusted when the dependence exist. Additionally, other issues such as sampling the extremes and the choice of threshold can also be problematic. Therefore, applying extreme value models is not always straightforward and the modification of traditional extreme value models needs to be considered to minimize the impact of these issues. These issues stimulated interest in examining the problems in applying extreme value modelling in finance, specifically with respect of risk measurements.

The physiological measurements of preterm infants often display instabilities because of their underdeveloped biological system. This induces the interest in examining the variability of these measurements since physicians believe that the variability supplies

information about underlying health states of preterm infants, over and above that provided by normal level of a physiological measurement solely. A suitable statistical model needs to be established in order to acquire an instantaneous index of variability. Further, we will also examine the tail behaviour of the variability to understand the risk level of instability to potentially signal "abnormal" condition of preterm infants.

1.2 Previous Research

Extreme value theory has been used to develop statistical models for rare events in many disciplines for the applied sciences over the last 50 years, (Pickands 1975; Danielsson and de Vries 1997; Coles 2001). The distinguishing feature of an extreme value analysis is that it assesses a data generating processes of rare events – in other words, tail behaviour. The extrapolation of tail behaviour is accomplished by the asymptotic extreme value theory (EVT). Extreme value theory supplies the asymptotic motivated approximate distributions in describing extremes, providing flexible and simple parametric models for fitting tail-related distributions. There are two very commonly used forms of such asymptotic motivated distributions. Under certain conditions, the distribution of a series of maxima/minima of a sequence of independent and identically distributed random variables is shown to converge to the Gumbel, Frechet, or Weibull distributions (Coles 2001). A unified form of these three distributions is called the generalized extreme value (GEV) distribution. Davison and Smith (1990) propose another asymptotic justified extreme model to describe the distribution of excess over a high threshold u, which is independent from the model. The model is named Generalized Pareto Distribution (GPD) and can minimize the problem of being wasteful of extreme information for gathering more extreme data compared to maxima over blocks as used for the GEV. Both GEV and GPD distribution can be used to estimate the very high quantiles (Danielsson and de Vries 1997) since the EVT related distributions supply a reliable way of examining tail behaviour even out of the scope of observations, which can be very useful in evaluating financial risk such as VaR.

Extreme value models are becoming more popular and also sophisticated. In recent years great efforts have been made to take into account dependence of extremes (Davison and Smith 1990; McNeil and Frey 2000; Ferro and Segers 2003), covariates (and potentially parametric and non-parametric functions of covariates) for non-stationary sequences (Smith 1989; Davison and Ramesh 2000b; Pauli and Coles 2001), threshold choice (Frigessi, Haug, and Rue 2002; Behrens, Lopes, and Gamerman 2004; Tancredi, Anderson, and O'Hagan 2006), and multivariate extremes (Coles and Tawn 1991; Coles and Tawn 1994; Heffernan and Tawn 2004).

There are several modifications off the classical extreme value models with respect to dependence of extremes. One approach is to decluster the dependent extremes and then to apply standard extreme value models on independent sequences, see Ferro and Segers (2003) for recent developments in statistical declustering algorithms. Peaks-over-threshold (POT) is one such declustering approach, see Davison and Smith (1990). Another way of accounting effects of dependence is to apply standard extreme value models, but to adjust the uncertainty estimates for dependence using techniques like block bootstrapping, see Buishand (1993). However, the substantial drawback of these approaches is that although they do account for the dependence in the inferences, they do not provide additional understanding of the dependence process itself.

Financial returns typically show clusters of observations as the well known ARCH (autoregressive conditional heteroscedastic) (Bollerslev, Chou, and Kroner 1992) and the general form GARCH process (Bollerslev 1986). Given the particular form of dependence of observations, a two stage model in financial literature, such as McNeil and Frey (2000) and Chan, Deng, Peng, and Xia (2007), address the dependence in applying extreme models in VaR. The method uses a GARCH type model to capture dependence of returns and obtains standardized innovations at the first stage. The extreme value model GPD on the approximately independent residuals is followed at the second stage. But the uncertainty estimation of such a two stage model is problematic. Bali and Weinbaum (2007) consider a related extension of the classical GEV, where they combine the GEV and GARCH model and explicitly model the dependence behaviour in allowing all three parameters of GEV to vary in time. It is always challenging to justify the form of extreme models and to estimate parameters due to the inherent sparsity of tail information. It is well known that the shape parameter is very difficult to estimate because it requires very large datasets to get reliable estimates, see Coles (2001). Further, it is unclear whether it is physically meaningful to allow the shape to vary rapidly over time, as it is driven by marketing trading behaviour.

Except for dependence of extremes, threshold selection of a GPD model can also be problematic in applications (Tancredi, Anderson, and O'Hagan 2006). Threshold selection must satisfy a balance between reliability of asymptotic approximation and sample variance of estimators. Traditionally, threshold value is chosen (fixed) using various graphical diagnostics (Guillou and Hall 2001; Coles 2001). The threshold is then treated as a known fixed constant in latter inferences. This approach suffers from concerns over subjectivity about threshold choice and lack of accounting for threshold uncertainty in inferences. A recently developed approach by Dupuis (1998) aims at reducing subjectivity and ensuring robustness. However, this method still requires some subjective assessment. Meanwhile, both approaches cannot capture the extra uncertainty

associated with threshold choice.

There has been much recent research in the development of mixture type models, which typically treat threshold as a model parameter to be estimated. Behrens, Lopes, and Gamerman (2004) use a truncated Gamma distribution for the bulk of the distribution, and a GPD above threshold. Tancredi, Anderson, and O'Hagan (2006) propose a less restrictive approach to overcome the lack of a natural model below threshold, by an unknown number of uniform distributions up to a suitable threshold and a GPD above that threshold. Frigessi, Haug, and Rue (2002) take a slightly different approach, using a dynamically weighted mixture model of a single GPD and a light-tailed distribution for the bulk of the distribution, with a smooth weight function for the transition between the two distributions. This approach avoids threshold choice, but replaces this problem with choice of transition function parameters. The mixture models no longer treat threshold to be fixed and thus inferences based on it, such as quantile estimation, automatically account for the uncertainty associated with threshold selection. All these models consider a mixture distribution with a GPD for the upper tail only and are not suitable for the application where a flexible form is needed for both the upper and lower tails, e.g. both being heavier than normal as often observed in financial returns.

The stochastic volatility and autoregressive conditional heteroscedasticity class of models has become widely established and successful in modelling return variance processes in financial time series. A variety of volatility models have been discussed by financial economists, such as the autoregressive conditional heteroscedasticity (ARCH) model of Engle (1982), the generalized ARCH (GARCH) model of Bollerslev (1986), the exponential GARCH (EGARCH) model of Nelson (1991), and the stochastic volatility (SV) models of Melino and Stuart (1990). These models provide latent volatility estimates of underlying financial returns, as volatility is not directly observable. With the recent move towards the use of high frequency data, there is interest in alternative volatility estimators – realized volatility (or integrated volatility) (French, Schwert, and Stambaugh 1987; Andersen, Bollerslev, Diebold, and Ebens 2001; McAleer and Medeiros 2008). Although volatility type models have been widely discussed by financial economists in the literature over the last two decades, this seems to be uncommon in medical applications since variability of physiological measurements of patients is of general interest. Magni and Bellazzi (2006) apply a stochastic model to extract the time course of such variability from the self-monitoring blood glucose time series in diabetic patients and combine the variability information with other analyses, to evaluate the adequacy of the therapeutic protocol and to highlight periods characterized by increasing glucose instability.

1.3 Thesis Objective

Risk assessment procedures are important for many research areas since risk can be found everywhere. The instabilities of preterm infants require a statistical model to measure the risk of being unstable using raw physiology measurements. The rapid increase of global economics and intergrated international financial markets, has also induced growth in financial crises, e.g. the current financial crisis starting in 2008. Risk management therefore appears more and more important for both investors and regulators. In this thesis, we apply the extreme value modelling in risk measurements with application in finance and neonatal study and focus on solving extreme value modelling in finance, such as dependence of extremes and threshold choice of GPD.

The instabilities of preterm infants suggest that the information contained in variability of their physiological measurements, besides considering the normal level of a physiological measurement, can supply additional insights to help doctors in making decisions about underlying health states of preterm infants. A suitable statistical model needs to be built in order to acquire an instantaneous index of variability. A commonly used volatility model in finance has therefore been tried and tested with the neonatal data to see whether it is appropriate in capturing the variability of preterm babies. Also, a method of indicating the upper quantile of the estimated latent variability needs to be proposed, which may potentially be used to distinguish normal and abnormal situations.

With respect to general dependence in financial data, dependence structures need to be addressed in extreme modelling to account for the impacts in inference. The existing approaches in dealing with dependent stationary process of extremes suffer from either being unable to describe specific dependence structures or having difficulty in counting uncertainty as a whole. Model identification is another difficulty for complex extreme value models. In this thesis, we intend to explicitly model the dependence behaviour of extremes for a stationary extreme time series within a single stage model to avoid the difficulty of the two stage method in accounting for all uncertainties of the whole model.

Threshold choice is another common problem of extreme value modelling considered in this thesis, particularly in respect of extreme value modelling in finance. The recent development of mixture type extreme value models, which treats the threshold as an extra model parameter to be estimated, can solve the relatively subjective threshold decision-making and automatically account for the uncertainty associated with threshold selection. But such models in financial application are generally not applicable since financial data tend to be heavy-tailed in both sides. This thesis therefore considers an extension of a one tail mixture model to a mixture having two tails to suit financial applications, where a flexible tail model is required for both the upper and low tails

simultaneously. Additionally, a new VaR method will be considered to deal with dependence, threshold selection and uncertainty estimation problems in VaR as an application of the two tail mixture extreme model.

Because of the typical problem of lacking observation of extremes, this thesis uses Bayesian inference in estimation of proposed extreme value models to take advantage of using extra prior information such as experts' suggestions.

1.4 Structure of Thesis

This thesis focuses on developing new statistical techniques in extreme value modelling on risk analyses in finance and neonatal studies. The research presented in the thesis involves variety econometric and statistical methods, including extreme value theory, volatility models and Bayesian. The relevant background, extreme value theory, and related applications in finance are reviewed in Chapter 2. At the start of each chapter, a detailed literature review of relevance to only the corresponding chapter is given.

Chapter 2 reviews the relevant background, extreme value theory and extreme value modelling in finance. Particularly, the chapter discusses general difficulties in applying extreme value models in finance, including the sampling of extremes, dependence of extremes and the threshold choice problems, which are the focuses of Chapter 4 to Chapter 6.

Chapter 3 considers the stochastic volatility model (SVM) fitted using Bayesian inference and a particle filter to capture the on-line latent volatility of oxygen concentration for preterm babies. Because instabilities are of concern, useful information about the health state of preterm infants can be expected to be contained in the variability of the physiological measurements. The study also conducts a comparison among alternative volatility to evaluate the performance of the SVM. The results suggest that the volatility of oxygen concentration level of preterm infants can be captured instantaneously by the stochastic volatility model with on-line estimation via a particle filter. The generalized pareto distribution (GPD) is then applied on estimated volatility series of preterm infants to examine the tail behaviour and to infer the extreme quantiles. As the volatility estimates are dependent, the peaks-over-threshold (POT) method is used to deal with the dependence of extremes. The extremal index is considered in the associated inference.

One of the challenges with extreme value methods application in finance is to account for the temporal dependence between the observations, for example the stylised fact that financial time series exhibits volatility clustering. Various approaches have been

proposed to capture the dependence. Commonly a two stage approach is taken, where the volatility dependence is removed using a volatility model like a GARCH (or one of it's many incarnations) followed by application of standard extreme value models to the approximately independent residual innovations. In chapter 4, we propose a novel one stage approach, which makes parameter estimation and accounting for the associated uncertainties more straightforward than the two stage approach. The location and scale parameters of the extreme value distribution are defined to have a conditional autoregressive heteroscedasticity process. Essentially, the model implements GARCH volatility via the extreme value model parameters. Bayesian inference is used and implemented via Markov chain Monte Carlo (MCMC), to permit all sources of uncertainty to be accounted for. The model is applied to both simulated and empirical data to demonstrate performance in extrapolating the extreme quantiles and quantifying the associated uncertainties.

A new mixture model is proposed in Chapter 5 to simultaneously model both tails using the GPD and to account for the uncertainty about thresholds in any inferences made. More specifically, we define a mixture distribution with three components: a GPD for losses tail, a GPD for gains tail and a certain distribution for the non-extreme data between the two GPD thresholds. The proposed two tail mixture GPD model can estimate both thresholds along with other parameter and therefore can account for all uncertainties, including that of threshold uncertainty. Clearly this extension provides a very flexible model for capturing all forms of tail behaviours, potentially allowing for asymmetry in the distribution of the two tails. The model is able to capture two sided tails distribution simultaneously, which is more applicable than a one sided tail mixture GPD in finance. Bayesian inference is used to account for all uncertainties and enables inclusion of expert prior information, potentially overcoming the inherent sparsity of extremal data. Simulations show the reliability and flexibility of the proposed mixture model, followed by VaR forecasting for capturing returns during the current financial crisis. A comparison with other approaches for extremes is also conducted, including: validating the threshold estimation by the robust estimation as (Dupuis 1998); the comparison in the extreme quantile estimation through different approaches; and asymmetry analysis.

Chapter 6 is an extension of Chapter 5. A new extreme value modelling approach in Value-at-Risk (VaR) is proposed, using the mixture extreme model (GNG) developed in Chapter 5 to overcome the key issues of determining the threshold and accounting for uncertainty due to threshold choice. As the dependence of financial returns, a two-stage approach (GARCH-GNG) is adopted: volatility estimation in defining the dependence structure followed by conditional extremal modelling of the independent innovations

via the prosed two tail mixture extreme model. We apply the approached method in forecasting VaR for the Citigroup and S&P100 during two financial crisis period, 2001 and 2008 to demonstrate the advantages of the model in estimating and forecasting VaR. Additionally, an alternative volatility – realized volatility, proposed mixture extreme model in estimating and forecasting VaR of IBM are shown to see the differences from the GARCH-GNG model. The empirical results show that realized volatility based method tend to give a faster changing VaR and a narrower uncertainty interval associated compared with the method using latent volatilities.

Chapter 7 summarizes the thesis and discuss the further relevant research areas.

1.5 Thesis Related Publications and Papers

The main results from Chapter 4 were pubilished by Zhao, Scarrott, Oxley, and Reale (2010b), to appear in Journal of Mathematics and Computers in Simulation in 2010. Some of the results in Chapter 6 and Chapter 5 will be published in Journal of Applied Financial Economics in 2010 as Zhao, Scarrott, Oxley, and Reale (2010a). Chapter 4 results have been presented at the Modelling and Simulation conference 2009 (Zhao, Oxley, Scarrott, and Reale 2009). Some results in Chapter 3 have been presented at the Modelling and Simulation conference 2007 (Zhao, Hou, Lee, and et al. 2007). The results presented in Zhao, Oxley, Scarrott, and Reale (2009) are from Chapter 4, and Zhao, Scarrott, Reale, and Oxley (2009a) presents the results from Chapter 5 & Chapter 6 which have been published as working paper in Economics Department of Canterbury University, alongside submission to the Journal of Econometrics by Zhao, Scarrott, Reale, and Oxley (2009b).

The statistical research in this thesis is my own and has not been submitted elsewhere for the award of a higher degree. I wrote all my own computer code with advanced mathematics and statistics package MATLAB.

Chapter 2

Extreme Value Modelling - A Review in Finance

2.1 Introduction

The objective of extreme value modeling is to quantify the outcomes of a stochastic process which have a small probability of occurring and even to extrapolate outside the scope of observations. Extreme value techniques have become widely used in the last 50 years in many disciplines such as the risk assessment of maximum loss in financial markets (McNeil and Frey 2000; Behrens, Lopes, and Gamerman 2004), the extreme levels of a river in hydrology (Coles and Tawn 1996; Tancredi, Anderson, and O'Hagan 2006), the largest claim in the actuarial analysis (Embrechts 1999). The extrapolation for the tail behavior is accomplished by the asymptotic extreme value theory. The extreme value theory (EVT) supplies the asymptotic justified distribution for extrapolating the underlying data generating process for these extremes providing a flexible and simple parametric model for capturing tail-related behaviours.

The study of extreme events in finance has long been important for investors and regulators. The current financial crisis has stimulated the interest in such extreme adverse movements of the market. In this thesis, extreme value modelling in the risk measurement in finance is of interest, specifically we are interested in estimating Value-at-Risk. In this chapter we review the extreme value theory and modelling with the focus in financial applications with the discussion of issues in applying extreme value modelling in finance.

The chapter is organized as follows. Section 2.2 reviews the EVT based distributions. Section 2.3 discuses the common estimation methods of extreme models. The common

issues in extreme value modelling are addressed in Section 2.4. An introduction of VaR with extreme modelling is given in Section 2.5. Section 2.6 summarizes the chapter.

2.2 Extreme Value Modelling

Classical extreme value theory (EVT) has been used to develop two commonly used extreme value distributions. First, the asymptotic distribution of a series of maxima (minima), under certain conditions, can be well approximated by the Gumbel, Frechet, or Weibull distributions (Coles 2001). A unified form of these three distributions is called the generalized extreme value (GEV) distribution. The second commonly used extreme value distribution concerns the excess over a given threshold, where one is interested in modelling the behaviour of the excess over(below) a high(low) threshold is reached. EVT shows that the limiting distribution is a generalized Pareto distribution (GPD) (Davison and Smith 1990; Hosking and Wallis 1987). Both extreme value distributions can be used to reliably estimate very high quantiles permitting examination of the tail behaviour even out of the scope of observations (Danielsson and de Vries 1997).

2.2.1 Generalized Extreme Value Distribution

Given a time series $(X_1, \ldots, X_n)$ consisting a sequence of independent and identically distributed (iid) random variables, the maximum $M_{x,n} = \max(X_1, \ldots, X_n)$ converges in law (weakly) to the following generalized extreme value distribution:

$$G_{ev}(x|\xi,\sigma,\mu) = \begin{cases} \exp\left\{-\left[1+\xi(\frac{x-\mu}{\sigma})\right]^{-1/\xi}\right\} & \xi \neq 0 \\ \exp\left\{-\exp(-\frac{x-\mu}{\sigma})\right\} & \xi = 0 \end{cases} \quad 1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0, \sigma > 0 \quad (2.1)$$

and the density of GEV is:

$$g_{ev}(x|\xi,\sigma,\mu) = \frac{1}{\sigma} G_{ev}(x|\xi,\sigma,\mu)\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-(1+1/\xi)} \quad 1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0 \quad (2.2)$$

where, σ is the scale parameter, μ is the location parameter and ξ is the shape parameter which indicates the tail behaviour. The Fréchet type tail corresponds to $\xi > 0$ and Weibull type tail corresponds to $\xi < 0$. The Gumbel type is interpreted as the limit as $\xi \to 0$ of Equation 4.1 with $\xi = 0$. The Fréchet distribution corresponds to heavy-tailed distributions and has commonly been found to be the most appropriate for the heavy tail of financial data. The asymptotic distribution of the maximum can be estimated without making any assumptions about the nature of the original distribution of the observations

since the asymptotic distribution of the maximum (if it exists) always belongs to one of these three distributions, whatever the original distribution.

When the minima $m_{x,n} = \min(X_1, \ldots, X_n)$ is of concern, the distribution function is as $G_{ev}(-x|\xi, \sigma, -\mu)$ as $\min(X_1, \ldots, X_n) = \max(-X_1, \ldots, -X_n)$.

The inverse of distribution function of GEV for the maxima, $G_{ev}^{-1}(1-p)$ represent the quantile of $1-p$, here p is the small probability (upper tail) as $P(x > x_p) = p$, which can be calculated as:

$$x_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left\{1 - [-\log(1-p)]^{-\xi}\right\} & \text{for } \xi \neq 0 \\ \mu - \sigma \log[-\log(1-p)] & \text{for } \xi = 0. \end{cases} \quad 1 + \xi\left(\frac{x-\mu}{\sigma}\right) > 0 \tag{2.3}$$

x_p is known as the return level with the return period of $1/p$. It can be interpreted as it will appear a extreme greater than the return level x_p once every $1/p$ period on average or as the mean waiting time between specific extremal events. In risk management of finance, x_p is known as the Value at Risk (VaR) to denote the maximum possible loss within a certain period $1/p$. For example, a maximum loss for a period of 30 days which is no more than R_p is to be assumed with a 5% risk of failure. This implies that a return level of R_p with a return period 585 days ($p = 0.0017083$) as solved for $P(L(R_p) \leq 30) = 1 - (1-p)^{30} = 0.05$, and $L(R_p)$ denotes the time of first failure which assumed a Bernoulli distribution.

2.2.2 The Excess Over Threshold Model - Generalized Pareto Distribution

Davison and Smith (1990) propose another asymptotic justified extreme value model to describe the distribution of the excess over a high threshold u, namely generalized pareto distribution (GPD). The GPD has the benefit of using more sample information for tail estimation, as compared to the GEV which considers the block maxima. It therefore can minimize the problem of being wasteful of extreme information for gathering more extreme data compare to GEV. Let X be a random iid variable of GPD and represent the extremes above the threshold selected, its distribution function is:

$$G_{pd}(x|\xi, \sigma, u) = \begin{cases} 1 - \left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi} & \xi \neq 0 \\ 1 - \exp\left(-\frac{x-u}{\sigma}\right) & \xi = 0 \end{cases} \quad \sigma > 0, \quad x \geq u, \quad 1 + \xi\left(\frac{x-u}{\sigma}\right) > 0 \tag{2.4}$$

and the density is as:

$$g_{pd}(x|\xi,\sigma,u) = \begin{cases} \frac{1}{\sigma}\left[1+\xi\left(\frac{x-u}{\sigma}\right)\right]^{-(1+1/\xi)} & \xi \neq 0 \\ \frac{1}{\sigma}\exp\left(-\frac{x-u}{\sigma}\right) & \xi = 0 \end{cases} \quad \sigma > 0, \quad x \geq u, \quad 1+\xi\left(\frac{x-u}{\sigma}\right) > 0 \tag{2.5}$$

where, ξ is the shape parameter, σ is the scale parameter and u is the threshold. There are three type of tail distributions associated with GPD regarding to the shape parameter value. A exponential decayed type tail correspond to $\xi = 0$, considered in the limit $\xi \to 0$. The excesses above the threshold has a slowly decaying tail and no upper bound if $\xi > 0$. The excesses distribution has an upper bound of the distribution if $\xi < 0$. Therefore, the shape parameter of GPD is dominant in determining the qualitative behavior of the tail.

For the extremes x below a low threshold u, the distribution of GPD is $G_{pd}(-x|\xi,\sigma,-u)$ conditional on $x < u$.

Similar to the GEV, the inverse of distribution function of GPD for the upper tail, $G_{pd}^{-1}(1-p)$ represent the quantile of $1-p$ for the excess over threshold, here p is the small probability as $P(x > x_p) = p$. Given that $x > u$, the conditional quantile or return level of x_p can be calculated as:

$$x_p = \begin{cases} u - \frac{\sigma}{\xi}(1-p^{-\xi}) & \text{for } \xi \neq 0 \\ u - \sigma\log(p) & \text{for } \xi = 0. \end{cases} \quad \sigma > 0, \quad x \geq u, \quad 1+\xi\left(\frac{x-u}{\sigma}\right) > 0 \tag{2.6}$$

Let $P(x > u) = p*$, the unconditional return level of q_p is given by:

$$q_p = \begin{cases} u - \frac{\sigma}{\xi}(1-(p/p^*)^{-\xi}) & \text{for } \xi \neq 0 \\ u - \sigma\log(p/p^*) & \text{for } \xi = 0. \end{cases} \tag{2.7}$$

In VaR, the risky quantile q_p implies the maximum loss (or maximum gain) R_p within $1/p$ period.

2.3 Estimation Methods

There are different estimation methods of extreme value models including graphical based method, moment based method, such as probability-weighted moment estimation and L-moments, maximum likelihood (ML)and Bayesian estimation. The ML method is the most commonly used approach, despite the likelihood being regular for only $\xi > -1/2$. L-moments can be preferred in finance studies since it can be used to get reliable

estimates for heavy tails. Bayesian inference has been well developed as it can use experts' knowledge through prior information to help the inherent sparsity of extreme observations. In this section, we review ML, L-moments and Bayesian methods only for brevity.

2.3.1 Maximum Likelihood Estimation

ML of GEV

The log-likelihood function for a sample $X_1, \ldots, X_m$ of i.i.d GEV random variables is given by

$$\log L(\xi, \sigma, \mu) = -m \log \sigma - (1+1/\xi) \sum_{i=1}^{m} \log \left[1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right] - \sum_{i=1}^{m} \left[1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right]^{-1/\xi} \tag{2.8}$$

with $1 + \xi \left(\frac{x_i - \mu}{\sigma}\right) > 0, \quad \xi \neq 0, \quad i = 1, \ldots, m.$

$$\log L(\sigma, \mu) = -m \log \sigma - \sum_{i=1}^{m} \left(\frac{x_i - \mu}{\sigma}\right) - \sum_{i=1}^{m} \exp\left[-\left(\frac{x_i - \mu}{\sigma}\right)\right], \quad \xi = 0, \quad i = 1, \ldots, m \tag{2.9}$$

Since the support of GEV depends on the unknown parameter values, the usual regularity conditions, underlying the asymptotic properties of maximum likelihood estimators, are not satisfied as studied in Smith (1985). In the case $\xi > -0.5$, the usual properties of consistency, asymptotic efficiency and asymptotic normality hold as $m \to \infty$, $\sqrt{m}\left((\hat{\xi}, \hat{\sigma}, \hat{\mu}) - (\xi, \sigma, \mu)\right) \xrightarrow{D} N(0, V), \xi > -0.5$, where V is the inverse of the Fisher information matrix. In applications, the sample with a very short tail as $\xi < -0.5$ is vary rare. As the majority of financial series have a positive tail index $\xi > 0$, ML offers a good tool for estimation of GEV. Follow from the approximate normality of the ML, a $100(1-\alpha)\%$ confidence interval for GEV parameters $\theta_i, i = 1, 2, 3$ is given by

$$\theta_i \pm \Phi^{-1}(1 - \alpha/2)\sqrt{\frac{\hat{v}_{i,i}}{m}}$$

where, $\hat{v}_{i,i}$ denotes the ith diagonal element of V after replacing the unknown parameters by their estimates. However, the confidence interval based on the normal limit results may be misleading as the normal approximation to the true sampling distribution of the respective estimator may be rather poor. In general, better approximations can be obtained by the profile likelihood function as Barndorff-Nielsen and Cox (1994). For

example, the profile likelihood confidence interval of ξ is given by

$$CI_\xi = \left\{ \xi : -2 \log \left(\frac{L_p(\xi)}{L_p(\hat{\xi})} \right) \leq \chi^2(1-\alpha) \right\} \tag{2.10}$$

where, $L_p(\xi) = \max_{\sigma,\mu|\xi}(L(\xi,\sigma,\mu))$ and the ratio of the profile likelihood follow a χ^2 distribution. The profile likelihood can also used to get the confidence interval for extreme quantile (return level or VaR) and its asymmetry feature is more appropriate for those with long return period as less information of the further apart extremes. This is particular preferred for financial application as most application with a relative heavy tail (with positive ξ). In the rest of the thesis, the confidence interval for extreme modelling with ML estimation are all based on profile likelihood approach.

ML of GPD

Given a value of threshold u, and the original sample $X_1, \ldots, X_n$, the extremes over u composite a new sample $X_1, \ldots, X_m$, $m < n$. The log-likelihood function for a GPD random variables is given by

$$\log L(\xi, \sigma|u) = -m \log \sigma - (1 + 1/\xi) \sum_{i=1}^{m} \log \left[1 + \xi \left(\frac{x_i - u}{\sigma} \right) \right] \tag{2.11}$$

with $1 + \xi \left(\frac{x_i - u}{\sigma} \right) > 0, \quad \xi \neq 0, \quad i = 1, \ldots, m.$

$$\log L(\sigma|u) = -m \log \sigma - \sum_{i=1}^{m} \left(\frac{x_i - u}{\sigma} \right), \quad \xi = 0, \quad i = 1, \ldots, m \tag{2.12}$$

Similar as GEV, the ML estimator hold the usual properties of consistency, asymptotic efficiency and asymptotic normality when $\xi > -0.5$. Again we prefer to use profile likelihood to get the confidence interval for both parameters and extreme quantiles interested when apply ML.

2.3.2 L-moments

L-moments of GEV

L-moments are certain linear combinations of probability weighted moments that have simple interpretations as measures of the location, dispersion and shape of the data sample. Let the random variable X has a distribution of GEV and the ordered sequence

$X_{n:n} \leq \cdots \leq X_{1:n}$, the first few L-moments are defined as:

$$\begin{aligned}
\lambda_1 &= E(X_{1,1}) \\
\lambda_2 &= \frac{1}{2}E(X_{1,2} - X_{2,2}) \\
\lambda_3 &= \frac{1}{3}E(X_{1,3} - 2X_{2,3} + X_{3,3}) \\
\lambda_4 &= \frac{1}{4}E(X_{1,4} - 3X_{2,4} + 3X_{3,4} + X_{4,4})
\end{aligned} \tag{2.13}$$

The first L-moment is the measure of location and the second L-moment is the measure of the dispersion of the data values about their mean. By dividing the higher-order L-moments by the dispersion measure, the L-moment ratios can be obtained as $\tau_r = \lambda_r/\lambda_2, \quad r = 3, 4 \ldots.$ τ_3 and τ_4 are the measure of the skewness and kurtosis of the distribution. For GEV distribution, the first two L-moments and moments ratio can be calculated as:

$$\begin{aligned}
\lambda_1 &= \mu - \sigma(1 - \Gamma(1 - \xi))/\xi \\
\lambda_2 &= -\sigma(1 - 2^{\xi})\Gamma(1 - \xi)/\xi \\
\tau_3 &= 2(1 - 3^{\xi})/(1 - 2^{\xi}) - 3 \\
\tau_4 &= (1 - 6 \times 2^{\xi} + 10 \times 3^{\xi} - 5 \times 4^{\xi})/(1 - 2^{\xi})
\end{aligned} \tag{2.14}$$

The skewness and kurtosis are independent from μ and σ and only depend on ξ. The shape parameter of GEV, ξ can be estimated by solving the equation $(1-3^{\hat{\xi}})/(1-2^{\xi}) = (\hat{\tau}_3 + 3)/2$. The equation can be solved using numerical method. The location μ and scale σ then can be estimated as:

$$\begin{aligned}
\hat{\sigma} &= \frac{-\hat{\lambda}_2\hat{\xi}}{(1 - 2^{\hat{\xi}})\Gamma(1 - \hat{\xi})} \\
\hat{\mu} &= \hat{\lambda}_1 + \frac{\hat{\sigma}\{1 - \Gamma(1 - \hat{\xi})\}}{\hat{\xi}}
\end{aligned} \tag{2.15}$$

L-moments is not sensitive to the outliers and is suitable for the application with heavy tails, which is the case of financial applications.

L-moments of GPD

If the random variable X (excess over threshold) follow a GPD ($\xi < 1$) and the first two L-moments can be calculated as:

$$\begin{aligned} \lambda_1 &= \frac{\sigma}{1-\xi} \\ \lambda_2 &= \frac{\sigma}{(1-\xi)(2-\xi)} \end{aligned} \tag{2.16}$$

And the L-moments estimates of GPD are as:

$$\begin{aligned} \hat{\sigma} &= (1-\hat{\xi})\bar{x} \\ \hat{\xi} &= 2 - \bar{x}/\hat{\lambda}_2 \end{aligned} \tag{2.17}$$

2.3.3 Bayesian Inference with Extremes

The typical problem with fitting extreme value models is the inherent lack of observations (Coles and Powell 1996). Therefore other information of the distribution can be useful if the estimation method can take account these information such as the expert knowledge. Bayesian inference can take advantage of using a prior distribution for the model parameters and therefore a good estimation tool for extreme value models supplementing the sparsity of extreme data. Also, the posterior distribution contains more information of the model parameter compared to ML since Bayesian analysis treats the parameter as a random variable and consider the uncertainty of the parameter in the inference. For example, when estimate the extreme quantile x_p (VaR or return level), the Bayesian inference can give the distribution of x_p including the uncertainty of the parameters.

There is no standard approach for defining the prior in extreme value models and it should vary application by application. Generally a relatively non-informative prior can be used when less information available and an relative informative prior otherwise. In particular, Coles and Tawn (1996) estimate the GPD via Bayesian inference with priors specified on the quantile differences since expert prior beliefs are generally easier to elicit on the quantiles themselves, rather than more directly on the parameters. The formulation of the prior elicited on the quantile differences also permits consideration of the known negative dependence between the shape ξ and scale σ parameters of the GPD. A gamma prior distribution is used to describe the quantile differences (details to be discussed in the relevant chapters). And it seems more recent research prefer this prior in applications, such as Coles and Pericchi (2003) and Behrens, Lopes, and Gamerman (2004).

2.4 Common Issues with Extreme Value Modelling

The typical problem with extreme value models is the inherent sparsity of extremal data and it can result in model classification and parameter estimation problems, particularly with a complex model structure. Bayesian inference is therefore more preferred in a complicate extreme model as it can take advantage of any expert prior information. In the thesis, we use Bayesian inference to estimate developed extreme value models.

We will also address other issues in the thesis including dependence of extremes, lacking of extremal observations and the choice of threshold. The followings section provides a general discussion of these issues and more details can be found in the related chapters.

2.4.1 Dependence of Extremes

The theory underlying the classical generalized extreme value (GEV) distribution gives the asymptotically justified limiting distribution of the maximum or minimum of a sequence of independent and identically distributed random variables. In most applications, particular the context of time series, the class of iid processes is too restrictive. In deed, the extremal events often tend to occur in clusters caused by the local dependence in the data.

Although under certain conditions (Beirlant, Goegebeur, Segers, and Teugels 2004), short range dependence between the observations can also lead to the distribution of the maximum or minimum remaining within the same GEV distribution family. But these conditions can not always be satisfied and the information contained in a extreme sequence with dependence is not as much as a iid sequence and the statistical inference need to be adjusted when the dependence exist. Therefore, it is straightforward to try reducing the dependence when sampling the extremes in accounting for the impact of the dependence. One of these approaches is to decluster the dependent extremes and then apply standard extreme models to the non-dependent sequence using statistical declustering algorithms, see Ferro and Segers (2003) for recent developments. The methods that are used to identify clusters define different estimators, the two most commonly used being blocks and runs declustering methods (Leadbetter, Weissman, de Haan, and Rootzen 1989). Hsing (1993) pointed out that a problem with these estimators is the selection of the declustering parameters, which is largely arbitrary, e.g. the choice of run length usually has a significant influence on the estimate of the cluster characteristic. Examples of other cluster characteristics are the cluster maximum, which is the focus of peaks-over-threshold modelling (Davison and Smith 1990), and the excess height statistic used by Leadbetter (1995).

The extremal index $\theta \in (0, 1]$ is involved in the inference of clusters. The extremal index is a quantity allows us to characterise the relationship between the dependence structure of the data and their extremal behaviour, see Leadbetter (1983). A strictly stationary iid process implies $\theta = 1$ and a smaller extremal index implies a strong dependence affect on extremal behaviour; if $\theta < 1$ then exceedence tend to cluster in the limit. To explain, we use an example to show the key function of extreme index in inference. Suppose we are interested in a VaR within next 50 trading days with a failure risk of 5%. Let the estimated extreme quantile 0.1% and 0.2% are -12% and -10% respectively based on classical GEV. The VaR is -12% as $(1 - 0.1\%)^{50} \approx 1 - 5\%$ if the return sequence is iid (or $\theta = 1$). The VaR is -10% as $(1 - 0.2\%)^{25} \approx 1 - 5\%$ with a non-iid sequence as $\theta = 0.5$.

The extremal index is one such characteristic, which has an interpretation due to Leadbetter (1983) as the reciprocal of the mean cluster size. Hsing (1987) shows that clusters of exceedence may be considered independent in the limit. Consequently, a common approach to inference is to identify independent clusters of exceedence above a high threshold instead. Estimation of the extremal index has been developed by Smith and Weissman (1994a), Weissman and Novak (1998) and Ferro and Segers (2003); for an application in finance see Longin (2000).

Another way of counting the affect of dependence is to apply standard extreme value models, but adjust the uncertainty estimates for the dependence using techniques like block bootstrapping, see Buishand (1993). However, the substantial drawback with both these approaches above is although they do account for the dependence in the inferences, they do not provide additional understanding the dependence process itself.

Financial returns typically show clusters of observations as well known ARCH (AutoRegressive Conditional Heteroscedastic) process and the general form GARCH process by Bollerslev (1986). Given the particular form of dependence of observations, a two stage model in the finance literature (McNeil and Frey 2000; Chan, Deng, Peng, and Xia 2007; Zhao, Scarrott, Oxley, and Reale 2010a) used addressing the dependence in applying extreme models. The dependence in the returns is captured in the first stage induced by the volatility clustering using a GARCH type model, followed by a second stage as extreme value modeling of the approximately independent residual innovations. But the uncertainty estimation of such two stage model is problematic.

2.4.2 The Choice of Threshold

The GPD distribution is conditional on $x > u$. The choice of a fixed threshold value is a key component for modelling using the GPD. However, threshold selection for the

GPD can also be problematic. The principle of threshold selection is to balance the reliability of the asymptotic approximation versus the sample variance of estimators as well. The threshold must be sufficiently high to ensure the threshold excesses have a corresponding approximate distribution within the domain of attraction of the generalized Pareto family. However, the threshold cannot be too high as this will reduce the sample information for inferences.

The standard practice is to adopt as low a threshold as possible, subject to the limit model providing a reasonable approximation (Coles 2001). Traditionally, two methods are used for this: one use various graphically diagnostics prior to model estimation; the other is to assess the features of model fitting for a range of potential thresholds. Specifically, the first method is based on the mean of GPD and to pre-extrapolate the exceedence over the threshold, $X - u$, via mean residual life plot (provided $\xi < 1$). The expected value of excess over threshold, for $(X - u) \sim G_{pd}(\xi, \sigma))$, is

$$E(X - u) = \frac{\sigma}{1 - \xi}$$

, provided $\xi < 1$. If the GPD is valid for excesses of a threshold u_0, it should be valid for all $u > u_0$ subject to the change of scale parameter σ_u. Hence, for $u > u_0$

$$\begin{aligned} E(X - u|X > u) &= \frac{\sigma_u}{1 - \xi} \\ &= \frac{\sigma_{u_0} + \xi u}{1 - \xi} \end{aligned} \tag{2.18}$$

So, the expected excess over threshold is a linear function of u and expected to change linearly with u, at which the GPD is appropriate.

The other method is to assess the threshold choice by the stability of the parameter estimators after fitting the model based on different choices. Above a level u_0, the asymptotic motivation for GPD is valid and the estimates of shape and scale parameters should be approximately constant.

However subjective assessment is required for above two methods, which can lead to different choices amongst scientists depending on their expertise and beliefs. Further, these approaches are not able to account for threshold uncertainty in inferences. Once a suitable value has been determined, the threshold is then treated as a known fixed constant in latter inferences.

A recently developed approach due to Dupuis (1998) aimed at reducing the subjectivity and ensure robustness. However, this method still requires some subjective assessment. For some applications, the threshold selection can be critical for the extrapolated tail behaviour, so the extra uncertainty associated with the threshold choice needs to be

accounted for. In estimating VaR, uncertainty estimation is also important for risk control.

The classical GPD is conditional on the realization that $u(x) = u$ for a prefixed threshold u. The lacking of general solution of threshold choice induces latent uncertainty about the threshold. Moreover, even the prefixed threshold may have good properties for the observed sample, it is not necessarily representative for other samples. Therefore, the variation of the threshold selecting is also need to be considered as it is part of the uncertainty of the model. There has been much recent research in the development of mixture type models, which typically treat the threshold as a model parameter to be estimated, and so also automatically accounts for the uncertainty associated with the threshold selection, see Behrens, Lopes, and Gamerman (2004), Tancredi, Anderson, and O'Hagan (2006) and Zhao, Scarrott, Oxley, and Reale (2010a). Frigessi, Haug, and Rue (2002) take a slightly different approach, using a dynamically weighted mixture model to avoid the threshold choice through a transition function between the extreme and bulk distributions.

2.5 Extreme in VaR

In the context of estimating Value-At-Risk (VaR), we are concerned, for example, with the possible extreme loss incurred in a day by the adverse market movement, see Duffie and Pan (1997). The prediction of the probability of the adverse market of a portfolio is key for risk management. Naturally, extreme value theory based models supply a statistically justifiable model to capture the tails of the underlying data generating process. The EVT based model can fit extreme quantiles better than other conventional approaches. Traditional extreme value approaches treat the upper (gains) and lower (losses) tail separately and does not require the symmetry of application. With the fact that most financial series are asymmetric and the EVT based models are advantageous over the models which assume symmetry such as normal distribution, t-distribution, GARCH type models with symmetry assumption.

There has been a number of extreme value studies in the financial literature recently. McNeil (1997) estimates the tails of loss severity distribution and McNeil (1998) calculates quantile risk measures for financial time series using extreme value theory. Embrechts, Klüppelberg, and Mikosch (1997) discusses the extreme value theory in finance and insurance comprehensively. (Embrechts 2000) examines the potentials and limitations of the extreme value theory in risk management. McNeil and Frey (2000) develop a two stage model in assessing the tail related risk for heteroscedastic financial time series.

Both block maxima (GEV) approach and excess over threshold (GPD) approach can be used in studying the extremal events. However, the GPD is more commonly used in estimating VaR as it can collect more tail information. Denoted the return sequence by $X_1, \ldots, X_n$ and the probability for $X < x$ as $F(x)$, the distribution function in estimating VaR for a quantile of q ($q > F(u)$) with GPD for excess over the threshold u can be written as:

$$F(x|\xi, \sigma, u) = \begin{cases} 1 - F(u)\left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi} & \xi \neq 0 \\ 1 - F(u)\exp\left(-\frac{x-u}{\sigma}\right) & \xi = 0 \end{cases} \quad 1 + \xi\left(\frac{x-u}{\sigma}\right) > 0 \tag{2.19}$$

where $F(u)$ represent the probability of $x \leq u$, which is commonly estimated as the empirical proportion of excess over threshold as N_u/n. Here, N_u is the number of return observations above the threshold value. The risk quantile can be calculated according to Equation 2.7 as:

$$\widehat{VaR}_q = u - \frac{\sigma}{\xi}\left\{1 - \left[\frac{n}{N_u}(1-q)\right]^{-\xi}\right\} \tag{2.20}$$

2.6 Summary

Extreme value modelling makes it possible to concentrate on the tail behaviour suited towards tail-related inference. For risk measurement in finance, such as VaR, extreme value based models are advantageous in reliable extrapolation to rare events as they exhibit flexible of the tail behaviours (short or fat tails). The typical problem in tail related inferences is the inherent lack of of extreme information, for which Bayesian inference provides a way to take advantage of prior expert information. Other common difficulties in applying univariate extreme value modelling include dependence of extremes, and threshold choice of GPD. We will focus on these issues in the following chapters with application to problems in finance.

Chapter 3

Volatility and Extreme Value Models in Neonatal Study

3.1 Introduction

Physiological measurements from preterm infants can display instabilities because of their under-developed biological systems. A combination of physiological measurements and dynamic variability analysis could lead to an improved diagnostic ability and to a better understanding of the healthy status (Galland, Hayman, Taylor, Bolton, Sayers, and Williams 2000). Besides considering the normal level of a physiological measurement, the variability of the measurement can supply independent information about the underlying state of health. Estimation of the variability of certain physiological measurements from preterm infants is therefore of clinical interest. When the desire is to acquire an instantaneous index of variability along with the measurement, the underlying variability has to be estimated with the aid of a suitable model. Although this seems to be uncommon in medical applications, it has been widely discussed by financial economists in the literature over the last two decades and is referred to as volatility models.

The stochastic volatility and autoregressive conditional heteroskedasticity class of models has become widely established and successful approaches to the modeling of the return variance process in financial time series. A variety of volatility models have been discussed by financial economists, such as the autoregressive conditional heteroskedastic (ARCH) model of Engle (1982), the generalized ARCH (GARCH) model of Bollerslev (1986), the exponential GARCH (EGARCH) model of Nelson (1991), and the stochastic volatility (SV) models of Melino and Stuart (1990). These models estimate the latent volatility of the underlying financial product returns, as the volatility is not directly ob-

servable. With the recent move towards the use of high frequency data, there is interest in alternative volatility estimators, such as realized volatility (or integrated volatility) (Fleming, Kirby, and Ostdiek 2003; Barndorff-Nielsen and Shephard 2002a; McAleer and Medeiros 2008). In particularly, French, Schwert, and Stambaugh (1987) use daily returns to estimate monthly volatilities. Blair (2001) explore the incremental volatility information of high-frequency (5-min) stock index returns. Andersen and Bollerslev (1998) show that realized volatility computed from high-frequency intraday returns is of observed and effectively an error-free volatility measure.

In medical research, there are very few applications of volatility models. Magni and Bellazzi (2006) apply a stochastic model to extract the variability from the self-monitoring blood glucose of diabetic patients and combine the variability information with other analysis to evaluate the adequacy of the therapeutic protocol and to highlight periods characterized by increasing glucose instability.

This chapter applies a stochastic volatility model (SVM) fitted using Bayesian inference to the oxygen concentration of preterm infants. A sequential Monte Carlo methodology, particle filtering, is used to deal with the difficulty in estimating the non-liner feature of the SVM. The basic idea is to use Monte Carlo methods to generate a large number of particles (sample points) to approximate the posterior distribution of the states (volatility) and coefficients of the model. The standard assumptions are that the distributions of the two noise processes in the model are known, and continuous distributions are approximated by discrete random measures, which are composed of weighted particles of the unknown states (volatility) and coefficients, with the weights computed using Bayes theory. In this study, the auxiliary sampling importance resampling (ASIR) filter, see Pitt and Shephard (1999) and Sanjeev, Maskell, and Gordon (2002), is used together with a kernel smoothing approach for estimating the unknown model parameters. Particle filters are particularly attractive for applications requiring on-line estimation.

An alternative Realized Volatility estimator, which is recently widely discussed in the finance area and regarded as an approximation of realized variance based on high frequency intra-period data (Christensen and Prabhala 1998; Andersen, Bollerslev, Diebold, and Labys 2001; Oomen 2005), is used as a benchmark to evaluate the performance of the latent volatility estimators based on SVM. As a comparison, we also use the EGARCH model on the same data and compare the relative performance of the two latent volatility models with realized volatility. The relative performance of the two latent volatility models is first evaluated using the R^2 measure from a linear regression analysis involving realized volatility and each of the latent volatility models at a relative lower frequency separately. Several other numerical measures of closeness between the latent volatility estimates and realized volatility are also computed. They all show the

SVM estimates to be closer to realized volatility than the EGARCH estimates. The results suggest that volatility at high frequency can be captured instantaneously for the physiological measurements by using the SVM.

To understand the distribution of unusually high variabilities of the physiological measurements from preterm infants can help to highlight the instability of these babies and potentially help doctors determine the healthy status of them. An extreme value distribution (generalized Pareto distribution or GPD for short), widely used in many areas, such as finance, hydrology and engineering, in modelling the tail related distributions, is used to give reliable approximation to the tail distribution of the estimated variabilities. We therefore apply GPD on the estimated stochastic volatility series of preterm babies to infer the extreme quantiles, which supply the information of the tail region and indicate the instability of preterm infants. Due to dependence of volatility estimates, peaks-over-threshold (POT) is used to deal with the dependence in extremes and the extremal index has been used to account for the dependence in the consequent inference.

Section 3.2 describes the stochastic volatility model (SVM) used in capturing the variability of preterm babies. An on-line sequential estimation method using a particle filter for the SVM is detailed in Section 3.3. The estimation results of stochastic volatility are showed in Section 3.4 and the performance of SVM with a particle filter is assessed in Section 3.5 in comparing alternative volatilities. Estimate of the upper quantiles of volatility are obtained using an extreme value distribution in Section 3.6 and the summary of the chapter follows in Section 3.7.

3.2 The Stochastic Volatility Model

In comparison to the GARCH family of volatility estimators and as its name "stochastic" suggesting, the stochastic volatility (SVM) introduces an additional innovation term in the conditional variance process. In the simplest framework, the series of returns, R_t; is modelled as the product of two stochastic processes. If the logarithm of volatility (in squared term as conditional variances), σ_t; is assumed to follow an AR(1) process, the corresponding model, known as the autoregressive SV (AR-SV(1)) model, is given by:

$$\begin{aligned} R_t &= u + \sigma_t \epsilon_t \\ \log(\sigma_t^2) &= \alpha_0 + \alpha_1 \log(\sigma_{t-1}^2) + \varphi_t \end{aligned} \tag{3.1}$$

Where the two innovation terms are independently distributed as $\epsilon_t \sim N(0,1)$ and $\varphi_t \sim N(0, v)$.

In this study, we use the model form as:

$$\begin{aligned} z_t &= u + \exp(x_t/2)\epsilon_t \\ x_t &= \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 \zeta_t \end{aligned} \tag{3.2}$$

where, the first equation is the measurement model with the measurement term - z_t defined as $\log(oxygen_t/oxygen_{t-1})$ and the second equation represents the state model with state term - x_t defined as $\log(\sigma_t^2)$ to ensure the positiveness of the conditional variance. The two innovation terms, ϵ_t and ζ_t, are independent and both follow a standard normal distribution with the additional coefficient α_2. Parameter α_0 and u are real numbers, the coefficient $|\alpha_1| < 1$ and α_2 is a positive scale factor. The model includes a drift in the measurements model as the oxygen level is expected to increase slightly when preterm babies are getting maturer and their biological systems are developing. In a short time period, the oxygen level of these preterm infants can stay the same level responding to a zero drift.

3.3 Particle Filter with SVM

Carnero and Ruiz (2001) show that the AR-SV(1) model outperforms the more popular GARCH (1,1) model to represent the empirical regularities often observed in financial time series. However, one of the most important limitations of SV models is that the distribution of R_t conditional on past observations up to time $t-1$ is unknown. Consequently, the exact likelihood function is difficult to evaluate. In order to compute it, the vector of the unobserved volatilities has to be integrated out of the joint probability distribution. Furthermore, introducing the additional innovation term in the conditional variance process increases the flexibility of the model but it also increases the difficulty in parameter estimation because each shock is now described by two innovations, ϵ_t and ε_t. Another difficulty with the SVM is the non-linear features of the model. In this paper, the SVM is estimated using a particle filter algorithm, which is based on approximation in the representation of the desired distributions by discrete random measures. The standard assumptions are that the distributions of the two noise processes in the model are known, and continuous distributions are approximated by discrete random measures, which are composed of weighted particles of the unknown states (volatility x_t) and coefficients, with the weights computed using Bayes theory. Particle filters are particularly attractive for applications requiring on-line estimation.

3.3.1 Algorithms of Particle Filter

Particle filtering algorithms use the following relationship between posterior and prior:

$$\underbrace{p(x_{t+1}|z_{1:t+1})}_{posterior} \propto \underbrace{p(z_{t+1}|x_{t+1})}_{Likelihood} \underbrace{\int p(x_{t+1}|x_t)p(x_t|z_{1:t})dx_t}_{prior} \tag{3.3}$$

In words, the posterior density at time $t+1$ is proportional to the likelihood function times the prior by Bayes's Theorem. In application in the SVM model, $p(x_{t+1}|z_{1:t+1})$ and $p(x_t|z_{1:t})$ are the posterior densities at time $t+1$ and t respectively. The likelihood function $p(z_{t+1}|x_{t+1})$ is derived from the measurement model of Equation 3.2 at time $t+1$, and $p(x_{t+1}|x_t)$ is derived from the state model of Equation 3.2 at time $t+1$.

In general, it is not possible to solve Equation 3.2 analytically. Particle filters use simulation to perform filtering. The posterior at time t is approximated using a set of particles $\{x_t^{(i)}\}_{i=1}^n$ and associated weights $\{w_t^{(i)}\}_{i=1}^n$, which sum to unity. The posterior distribution is approximated by a discrete probability mass function whose support is the set of particles, and which has probability $w_t^{(i)}$ assigned to the i^{th} particle value. Using $\widehat{\pi}_t(x_t)$ to denote the particle filter approximation to the posterior at time t,

$$\widehat{\pi}_t(x_t) = \sum_{i=1}^{n} w_t^{(i)} \delta(x_t = x_t^{(i)}) \approx p(x_t|z_{1:t}) \tag{3.4}$$

where, δ is the Dirac delta function. The approximation for $p(x_{t+1}|z_{1:t+1})$ is as follows:

$$\widehat{\pi_{t+1}}(x_{t+1}) = \sum_{i=1}^{n} w_t^{(i)} p(x_{t+1}|x_t^{(i)}) p(z_{t+1}|x_{t+1}) \tag{3.5}$$

The particle filter can be viewed as two parts: Initiation and Iteration. Initiation produces a particle approximation to the prior $p(x_0) \equiv p(x_0|z_0)$. In this paper the initialization of the parameters at the starting time is gained by Gibbs Sampler; Iteration uses a particle approximation for $P(x_t|z_{1:t})$ at time t, $\{x_t^{(i)}, w_t^{(i)}\}_{i=1}^n$, to obtain a new particle approximation for $P(x_{t+1}|z_{1:t+1})$, which is $\{x_{t+1}^{(i)}, w_{t+1}^{(i)}\}_{i=1}^n$ at time $t+1$, when a new measurement z_{t+1} becomes available. The iteration step is split into propagation, re-weighting and re-sampling steps.

Importance sampling is used to obtain the particles and associated weights. If we use $q(x_{t+1}|z_{1:t+1})$ to denote the importance density that could be used to sample the particles

$\{x_{t+1}^{(i)}\}_{i=1}^{n}$ directly, then the weights are defined as:

$$w_{t+1}^{(i)} \propto \frac{p(x_{t+1}^{(i)}|z_{1:t+1})}{q(x_{t+1}^{(i)}|z_{1:t+1})} \tag{3.6}$$

From Equation 3.5, we could derive the following relationship:

$$p(x_{t+1}^{(i)}|z_{1:t+1}) \propto w_t^{(i)} p(x_{t+1}^{(i)}|x_t^{(i)}) p(z_{t+1}|x_{t+1}^{(i)}) \tag{3.7}$$

Thus, the weight could be updated as:

$$w_{t+1}^{(i)} \propto w_t^{(i)} \frac{p(x_{t+1}^{(i)}|x_t^{(i)}) p(z_{t+1}|x_{t+1}^{(i)})}{q(x_{t+1}^{(i)}|z_{1:t+1})} \tag{3.8}$$

If the importance density is only dependent on the current measurement $z_t + 1$, which is the SVM case, then Equation 3.9 could be simplified further as:

$$w_{t+1}^{(i)} \propto w_t^{(i)} \frac{p(x_{t+1}^{(i)}|x_t^{(i)}) p(z_{t+1}|x_{t+1}^{(i)})}{q(x_{t+1}^{(i)}|z_{t+1})} \tag{3.9}$$

3.3.2 Implementation Issues with Particle Filter

To implement a particle filter for the SVM, several issues have to be resolved such as re-sampling, choice of proposal density and kernel smoothing of unknown parameters.

Re-sampling

A degeneracy problem can arise after a number of time steps when only one particle has significant weight. Doucet (1998) showed that the variance of the importance weights increases over time, making degeneracy unavoidable. Thus, considerable computational effort is expended on updating particles whose contribution to the approximation of $p(x_{t+1}|z_{1:t+1})$ is negligible.

The basic idea of re-sampling is to eliminate particles that have small weights and to concentrate on particles with large weights. A new particle set is thus generated by sampling with replacement from the original set $\{x_{t+1}^{(i)}\}_{i=1}^{n}$ with probability $p(x_{t+1}^{(j)} = x_{t+1}^{(i)}) = w_{t+1}^{(i)}$. Here j is the particle index after re-sampling. The "updated relationship" is denoted as $parent(j) = particle(i)$. The weights are re-set to $1/n$ as the particles are independent and identically distributed from a discrete distribution.

There are a lot of ways of doing the re-sampling, such as systematic re-sampling and

re-sampling using standard Monte Carlo. In this paper, re-sampling with random quasi monte carlo (QMC) is used. The advantage of Random QMC is discussed by Fearnhead (2005).

Simply, a random QMC sequence $y_1, y_2, \ldots, y_n$ on $[0, 1]$ satisfies:

1. Marginally $y_i \sim Uniform(0, 1)$ for all $i = 1, \ldots, n$.
2. Jointly, $y_1, y_2, \ldots, y_n$ is a QMC sequence with probability 1.

In our one dimensional case, a simple random QMC sequence is obtained by letting the ordered sequence of points $\{y_i\}_{i=1}^n$ satisfy $y_i \sim Uniform(\frac{i-1}{n}, \frac{i}{n})$. For the particle filter that is used in this paper, re-sampling occurs within the propagation step.

Choice of Proposal Density

Different choices of proposal density can be made. A reliable importance density was proposed by Pitt and Shephard (1999): $q(x_{t+1}^{(i)}) \propto p(z_{t+1}|\mu_{t+1}^{(i)})p(x_{t+1}^{(i)}|x_t^{(i)})w_t^{(i)}$, where $\mu_{t+1}^{(i)}$ is defined as the mean of $p(x_{t+1}|x_t^{(i)})$, which is from the state model as Equation 3.2. Also considering the re-sampling stage above, the particle $x_{t+1}^{(j)}$ is assigned a weight as follows:

$$w_{t+1}^{(j)} \propto \frac{p(z_{t+1}|x_{t+1}^{(j)})}{p(z_{t+1}|\mu_{t+1}^{(parent(j))})} \tag{3.10}$$

Filters with this importance density and re-sampling stage are known as auxiliary sampling importance re-sampling (ASIR) filters.

Kernel Smoothing for Unknown Parameters

Section 2.2 focused on the estimation for the state term x_t. When unknown model parameters are present as well, Bayes's rule gives the following joint posterior distribution:

$$p(x_{t+1}, \theta_{t+1}|z_{1:t+1}) \propto p(z_{t+1}|x_{t+1}, \theta_{t+1})p(x_{t+1}, \theta_{t+1}|z_{1:t})p(\theta_{t+1}|z_{1:t}) \tag{3.11}$$

where θ is the vector of model parameters. Joint state and parameter estimation is achieved through the augmentation of the state space with the parameter vector. A Gaussian random walk for the parameters can then be specified to enable their adaption to a new measurement: $\theta_{t+1} = \theta_t + \eta_{t+1}$, where $\eta_{t+1} \sim G(0, M_{t+1})$ with pre-defined covariance matrix, M_{t+1}. However, the random walk implies an increase in the covariance, resulting in posteriors more diffuse than the actual ones. A natural approach to reducing the covariance is to use kernel smoothing with smoothing factor, $0 < h < 1$ as noted

by Liu and West (2001). The approximation of the distribution of the parameters by a mixture density is:

$$p(\theta_{t+1}|z_{1:t}) \approx \sum_{i=1}^{n} w_t^{(i)} G(\theta_{t+1}|m_t^{(i)}, h^2 V_t) \tag{3.12}$$

Here, the kernel locations $m_t^{(i)}$ is specified by a shrinkage rule that forces the particles to be closer to their mean:

$$m_t^{(i)} = (\sqrt{1-h^2})\theta_t^{(i)} + (1-\sqrt{1-h^2})\bar{\theta}_t \tag{3.13}$$

In the two equations above, $\bar{\theta}_t$ and V_t are the Monte Carlo mean and covariance matrices computed from all the particles with weights, $\{\theta_t^{(i)}, w_t^{(i)}\}_{i=1}^n$. Thus at each time step, particles for θ_{t+1} are drawn from the mixture density above. The kernel smoothing provides a flexible approach to deal with unknown parameters. Since the parameters are assumed to be fixed, we let $h = n^{-1/(4+d)}$ where $d = 3$, i.e. $h = n^{-1/7}$.

3.4 Data and SVM Estimates

3.4.1 Data

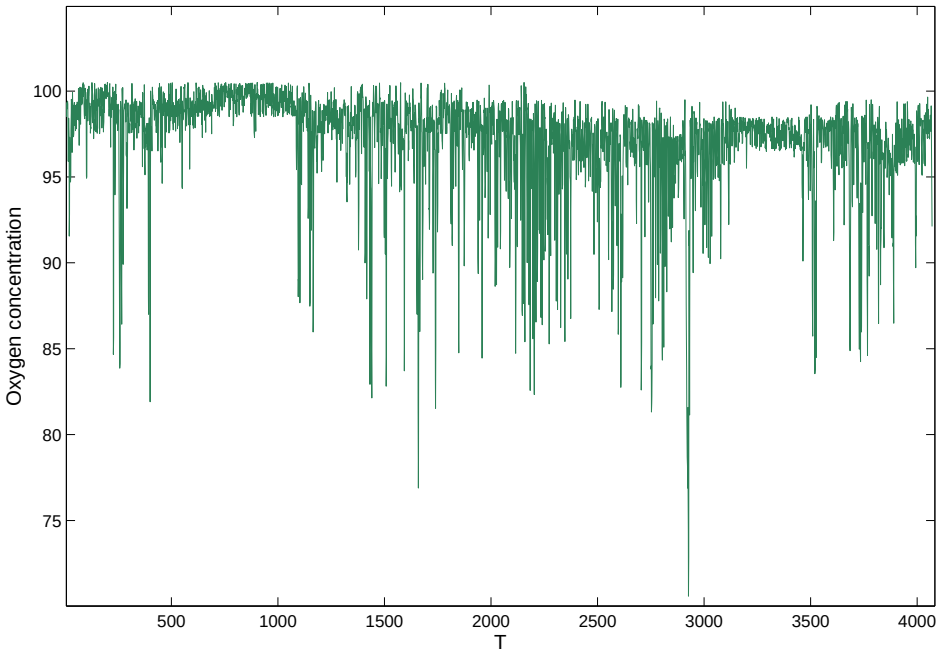

Figure 3.1: The sequence of Oxygen concentration measurements

The oxygen dataset, comprising measurements every 2 seconds, was supplied by Christchurch Women's Hospital and was originally collected for an earlier study entitled,

"Normal variation in oxygen levels in preterm babies". A segment of 4071 observations, shown in Figure 3.1. The variable z_t in the SVM is calculated as $\log(oxygen_t/oxygen_{t-1})$ and leads to a series of 4070 as shown in Figure 3.2 with clear clusters.

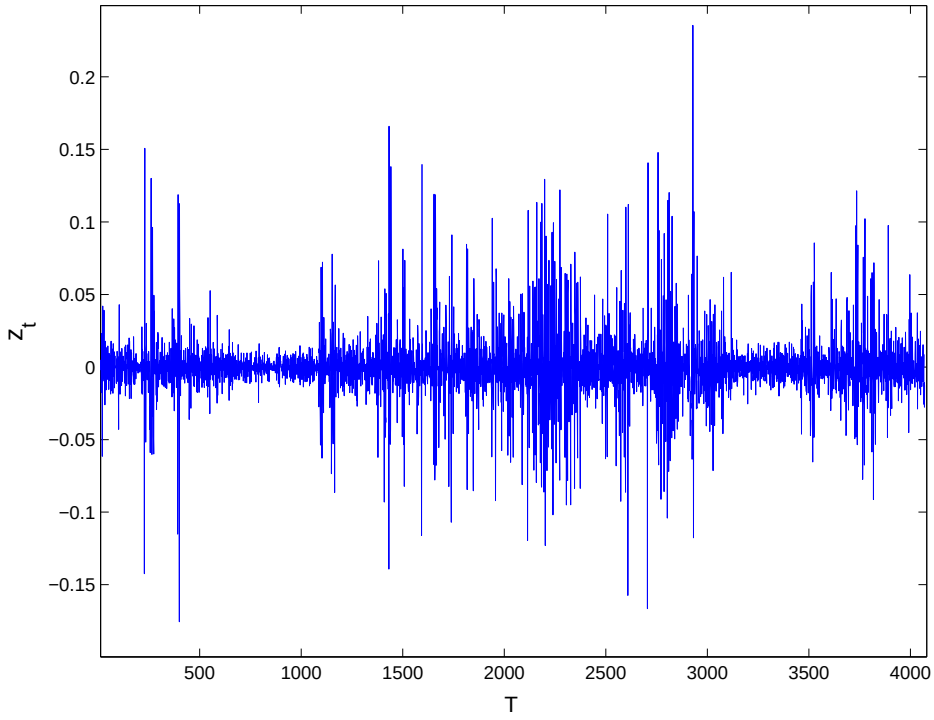

Figure 3.2: The sequence of z_t ($\log(oxygen_t/oxygen_{t-1})$).

3.4.2 Stochastic Volatility Estimates

Using the dataset above, the conditional volatilities σ_t at 2-second frequency are estimated based on the stochastic volatility model defined as Equation 3.2. Figure 3.3 shows the $x_t = \log(\sigma^2)$ time series estimated under SVM using a particle filter and the standardized innovation term as z_t/σ_t. There are clear volatility clusters in the estimates and the cluster feature the standardized innovation disappears after removing the volatility effects.

3.5 Performance of SVM

As demonstrated by Andersen and Bollerslev (1998), Barndorff-Nielsen and Shephard (2002c) and Andersen, Bollerslev, Diebold, and Ebens (2001), the sum of squared intraday returns is an approximation to the daily realized variance. This is named realized volatility (RV) and regarded as an estimate of integrated volatility. Unlike latent volatility, RV is model-free and based on the information within a time window rather than

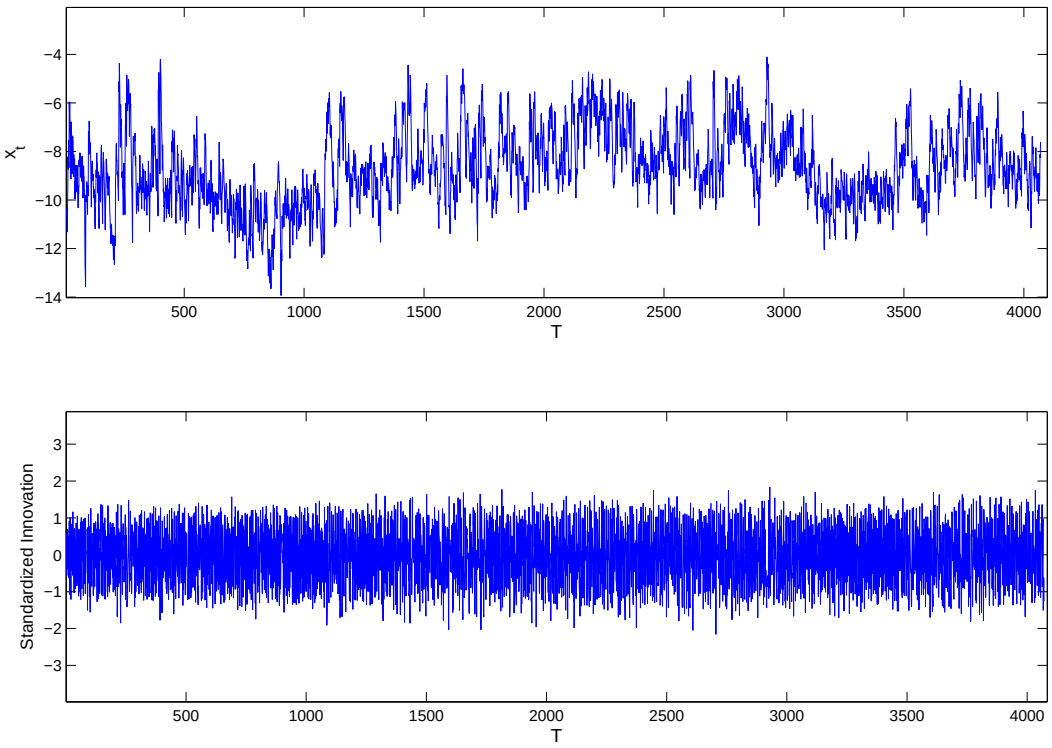

Figure 3.3: Time series plot of $x_t = \log(\sigma^2)$ estimates for oxygen measurements based on SVM model (Upper figure). The standardized innovation term z_t/σ_t (Bottom figure).

all previous information until time t. RV can be built up easily as an integrated measurement of the intra-day returns and is a consistent estimator of observed volatility when high frequency data are available without structural bias and serial correlation (Andersen, Bollerslev, Diebold, and Ebens 2001). Forecasting and modeling volatility is equivalent to forecasting and modeling realized volatility (Andersen, Bollerslev, and Meddahi 2005).

However, the calculation of RV requires a large amount of higher frequency data and may not always be available in applications. In this study, the aim is to quantify instantaneous variability of oxygen contestation and the instantaneous model free volatility is not applicable. In that case, we can only stay with latent volatility and RV is calculated at a relative low frequency rate and used as a benchmark to evaluate the performance of SVM volatilities estimated with a particle filter.

Besides volatility estimates under SVM, another commonly used latent volatility model is the "GARCH" family of volatility models. Following the introduction of ARCH processes by Engle (1982) and their generalization by Bollerslev (1986), there have been numerous modifications of this approach of modeling conditional volatility. Nelson (1991) proposed the exponential GARCH (EGARCH) model, which allows for asymmetric effects between positive and negative values. We use the EGARCH model in the study

for comparison with SVM since both the SVM and EGARCH use $\log(\sigma^2)$ instead of σ^2 to ensure positiveness of the conditional variance.

3.5.1 Alternative Volatility Estimators

EGARCH

The following EGARCH process is applied to oxygen concentration data:

$$\begin{aligned} z_t &= u + \exp(x_t/2)\delta_t \\ x_t &= \beta_0 + \beta_1 x_{t-1} + \beta_2[|\delta_{t-1}| - E(|\delta_{t-1}|)] + \beta_3\delta_{t-1} \end{aligned} \tag{3.14}$$

As with the SVM, z_t is the measurement term defined as $\log(oxygen_t/oxygen_{t-1})$ and δ_t serves as the shock from the mean. The x_t is defined as $\log(\sigma^2)$ of the z_t; δ_t is a Gaussian iid sequence and $E(|\delta_t|) = \sqrt{2/\pi}$. Maximum likelihood estimation is used to estimate the parameters and obtain the volatility σ_t sequence at every 2 seconds frequency.

Realized Volatility

Although the integrated variance of realized volatility can be regarded as an observed volatility, but it can only be obtained with high frequency data. However physicians are interested in the instantaneous variability of physiological measurements of preterm babies, we therefore can only calculate RV based on a relative lower time frequency and use it to assess consistency with the latent volatility estimates. The RV is obtained as the integrated volatility for every minute using the 2-second data, giving 135 volatility estimates. We also looked at different time-aggregation to check the robustness of the estimated realized volatility. Unlike financial data, the physiological measurements do not have microstructure bias but still have slight serial correlation problems for intra-period data.

3.5.2 Comparison of Alternative Volatilities

In order to compare the volatility estimators under SVM and EGARGH models with RV, the cumulative volatility per minute is calculated based on the estimated latent volatility per 2 second for both latent volatility estimates to standardize the time frequency with RV. Although the integrated volatility will reduce the differences of different volatilities, it will provide a overall indication of the closeness of latent volatility to realized volatility. Table 3.1 gives the summary statistics of the three volatility sequences at 1 minute

Table 3.1: Summary of different volatility estimates

Different Volatilities	σ(2 seconds)		σ(1 minute)		
	SVM	EGARCH	SVM	EGARCH	RV
Maximum	0.1685	0.1923	0.3234	0.4148	0.3297
Minimum	0.0009	0.0031	0.0189	0.0285	0.0229
Mean	0.0183	0.019	0.1139	0.1117	0.1071
Std.Dev	0.0163	0.0134	0.0716	0.0616	0.0709

frequency.

The time series of volatilities $\{\sigma_t\}$ from both latent volatility models are shown in Figure 3.4. Although the two sequences of estimates appear to be consistent over the entire time series, there are clear differences particularly for low and high volatilities.

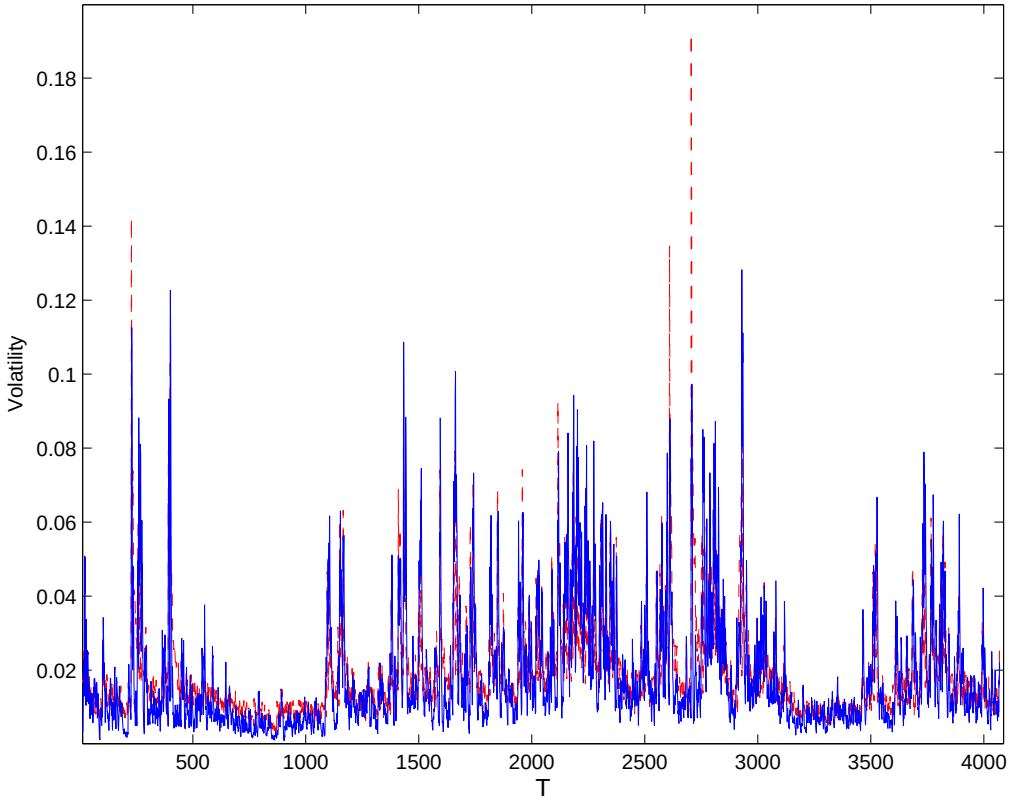

Figure 3.4: Latent volatility estimates. The solid line denotes the SVM volatility every 2 seconds and dashed line represents the EGARCH volatility every 2 seconds.

In order to compare the performance of the two latent volatility models, we use RV as a benchmark and exam the closeness of latent volatility and realized volatility. The proportion of the total variation of RV explained by the estimated latent volatility is

Table 3.2: Performance of alternative volatility models

	MAE	RMSE	HMAE	HRMSE
EGARCH	0.0206	0.0309	0.1829	0.225
SVM	0.0116	0.0171	0.1173	0.1486

used and measured using R^2 from a least-squares linear regression analysis as:

$$\begin{aligned} \sigma_{(RV,t)} &= a_0 + a_1\sigma_{(SVM,t)} + a_2\epsilon_t \\ \sigma_{(RV,t)} &= b_0 + b_1\sigma_{(EGARCH,t)} + b_2\epsilon_t \end{aligned} \tag{3.15}$$

The R^2 value for SVM volatility is about 93.95%, while for EGARCH, it is about 81.36%. This suggests that SVM volatility describes the dynamics of RV better than EGARCH volatility.

In addition to R^2, we compute the mean absolute error (MAE) and root mean squared error (RMSE), as well as the heteroscedasticity-adjusted mean absolute error (HMAE) and root mean squared error (HRMSE), to measure how far the estimated latent volatilities are from the realized volatility. These measures are defined as follows:

$$\begin{aligned} MAE &= \frac{1}{n}\sum_{t=1}^{n}\left|\sigma_{(latent,t)} - \sigma_{(realized,t)}\right| \\ RMSE &= \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\sigma_{(latent,t)} - \sigma_{(realized,t)}\right)^2} \\ HMAE &= \frac{1}{n}\sum_{t=1}^{n}\left|1 - \frac{\sigma_{(realized,t)}}{\sigma_{(latent,t)}}\right| \\ HRMSE &= \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(1 - \frac{\sigma_{(realized,t)}}{\sigma_{(latent,t)}}\right)^2} \end{aligned} \tag{3.16}$$

These measures are reported in Table 3.2. It is clear that both the MAE and RMSE of SVM are smaller than EGARCH. After the deviation between estimated latent volatility and realized volatility is adjusted for heteroscedasticity, SVM still outperforms EGARCH with smaller HMAE and HRMSE.

Further useful insight can be gain from Figure 3.5, which presents volatility per minute for the SVM, EGARCH and RV together. The SVM volatility estimates are of closer to the RV when compared to the EGARCH volatility estimates.

We further obtain approximate 95% confidence intervals for RV at every minute by using a simple bootstrap method, in that bootstrap creates 1000 bootstrap sample by

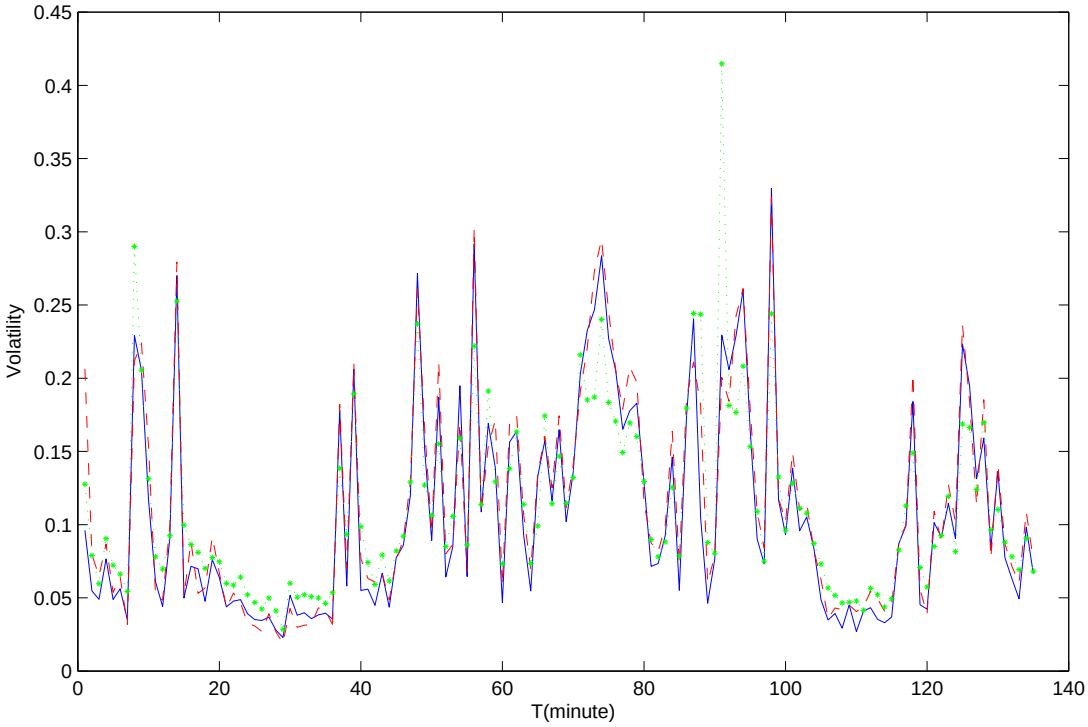

Figure 3.5: Latent volatilities and realized volatilities. The solid line denotes the RV volatility, dashed line represents the SVM volatility and dotted line with dot marker indicates the EGARCH volatility.

sampling with replacement from the non-scalar data arguments, to assess how close SVM volatility estimates are to RV. In Figure 3.6, the dashed line with circles is the pointwise confidence interval for each time-point. The star symbol represents realized volatility every minute and the dashed line with plus sign is the SVM volatility per minute. There are few points outside the confidence intervals, and these occur within the first ten time-points, corresponding to the initial period where the particle filter has not yet converged. The estimates from the particle filter improve as more measurements are processed, and so the estimates are closer together as time progresses. HMAE and HRMSE are 0.1187 and 0.1467 for the first half of the data and smaller HMAE and HRMSE are 0.1097 and 0.1375 for the second half of the data.

3.6 Upper Quantile of Stochastic Volatility

In order to understand the distribution of unusually high volatilities, so that potential instability of the physiological measurements can be highlighted to potentially help doctors determine the health status of preterm babies, we use an extreme value model applied to the estimated stochastic volatilities above. Specifically, the Generalized Pareto Distribution (GPD) for exceedences of a high threshold is used in extrapolating the upper tail

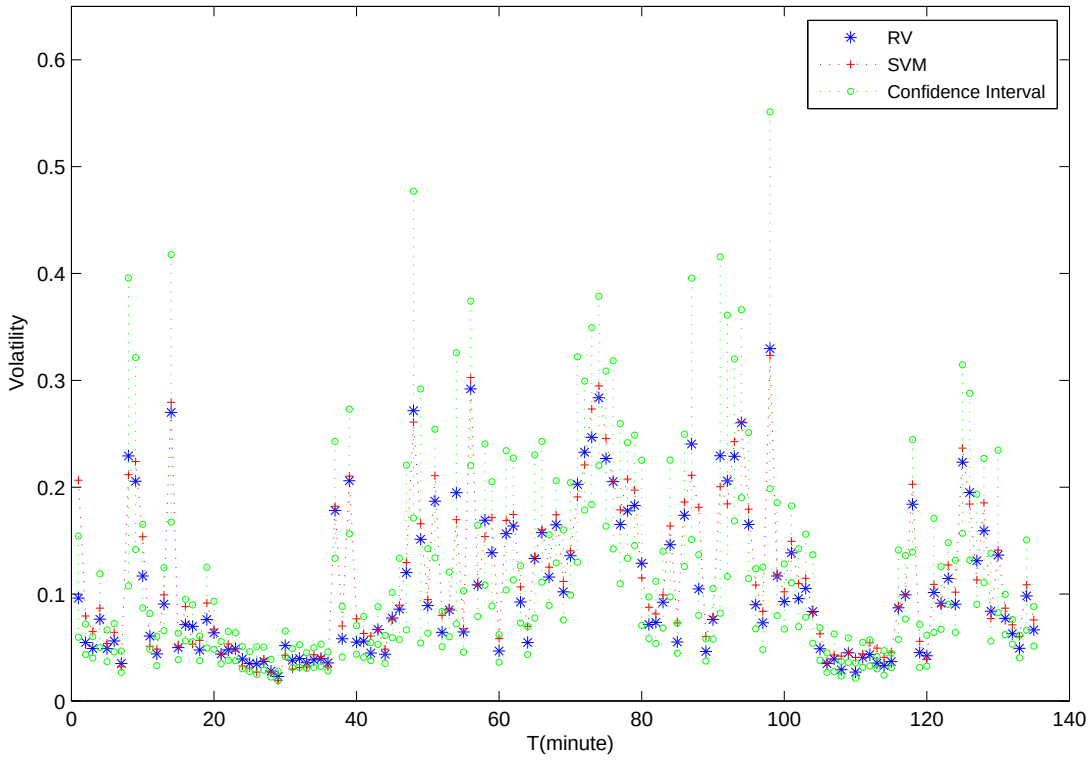

Figure 3.6: SVM and RV with CI. The "*" denotes the RV and "+" indicate SVM volatility every minute. The vertical dashed line give the 95% confidence interval of RV by bootstrap.

of the volatilities. For the stochastic volatility sequence $\{sv_t\}$, the distribution function of GPD is:

$$F(sv|\xi,\phi,u) = \begin{cases} 1-\eta_u\left[1+\xi\left(\frac{sv-u}{\phi}\right)\right]^{-1/\xi} & \xi \neq 0 \\ 1-\eta_u \exp\left(-\frac{sv-u}{\phi}\right) & \xi = 0 \end{cases} \tag{3.17}$$

where, ξ and ϕ ($\phi > 0$) are the shape and scale parameter of GPD, $\eta_u = P(sv_t > u)$ is the probability of exceeding the threshold u, and $1+\xi\left(\frac{sv-u}{\phi}\right) > 0$. The two parameter (ξ,ϕ) of GPD is estimated by maximum likelihood (ML) and the maximum likelihood estimates (MLE) of η_u is simply the sample proportion above u.

There are two issues in examining the upper tail distribution of the estimated stochastic volatility by GPD model. Firstly, the estimated sv is not an independent sequence since it is obtain by an autoregressive process in the SVM (AR(1)-SV). The classical GPD assumptions are that the underlying variables are from an iid process. One commonly used approach dealing with the dependence of extremes is to decluster the dependent extremes and then apply standard extreme models to non-dependent sequence in using statistical declustering algorithms, see (Ferro and Segers 2003) for recent developments.

Peaks-over-threshold (POT) is one such declustering approach, see (Davison and Smith 1990). In this section, POT method is used to reduce the dependence in sampling extremes. The estimated distribution contains the information of independent clusters rather than individual observations and the extremal index is used in the quantile estimation for the upper tail of estimated *sv*. In particular, the method by Ferro and Segers (2003) is used in the estimation of extremal index, see Section 2.4 for the literatures relevant.

Figure 3.7 shows the exceedence over a high threshold of estimated *sv* and these exceedences tend to appear in clusters (pair for this particular dataset). To account for the impact of the dependence, declustering the dependent extremes approaches is used and extremal index in the inference of clusters is estimated as Ferro and Segers (2003). In particular, the extremal index is used in estimating the quantile of upper tail of the estimated *sv*.

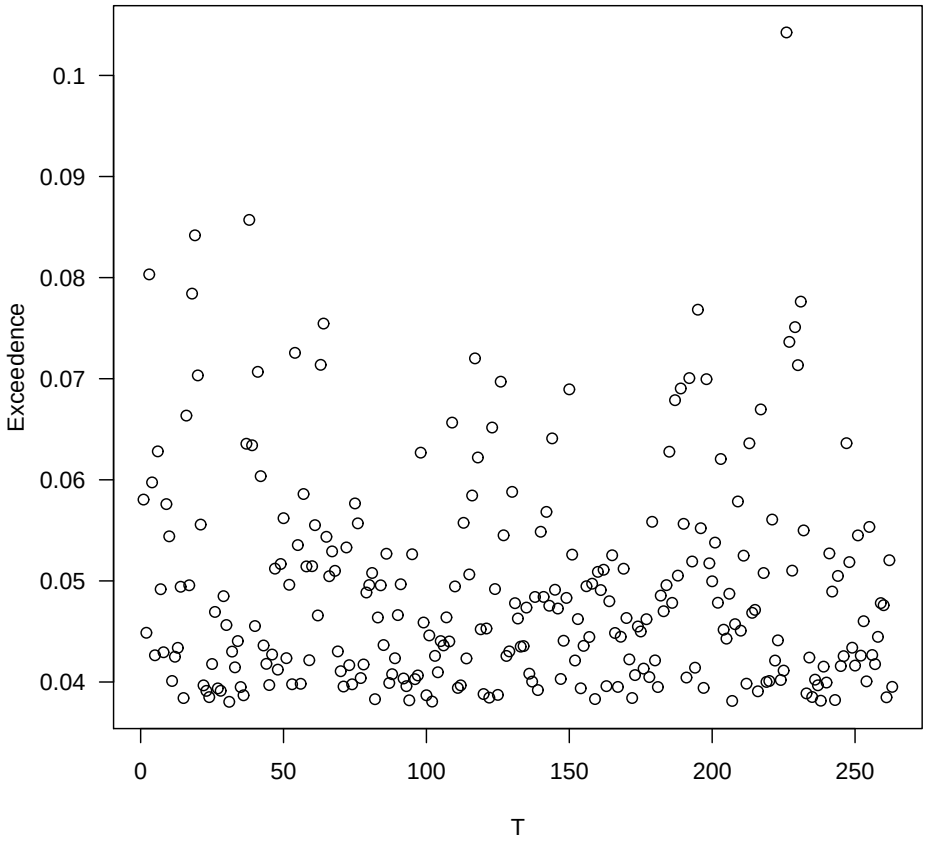

Figure 3.7: Exceedence over a high threshold

The second issue is threshold choice, which is analogous to bandwidth choice in kernel

density estimation, where the choice aims to balance between bias and variance. Too high a threshold will lead to a small dataset for parameter estimation and consequently a high variance of the model, whereas too low a threshold will lead to the asymptotic underlying the GPD approximation potentially being invalid leading to bias. In this chapter we will not discuses the uncertainty with threshold choice and treat it as a prefixed value using classical exploratory technique in examining the affect of threshold choice in the approximation to the excess distribution and the stability of model parameter estimation. Figure 3.8 gives the mean excess against the threshold choices for the estimated *sv*, and shows the linearity around $u = 0.038$. The parameter estimation against threshold in Figure 3.9 shows the stability in the model parameter estimation with the choice $u = 0.038$.

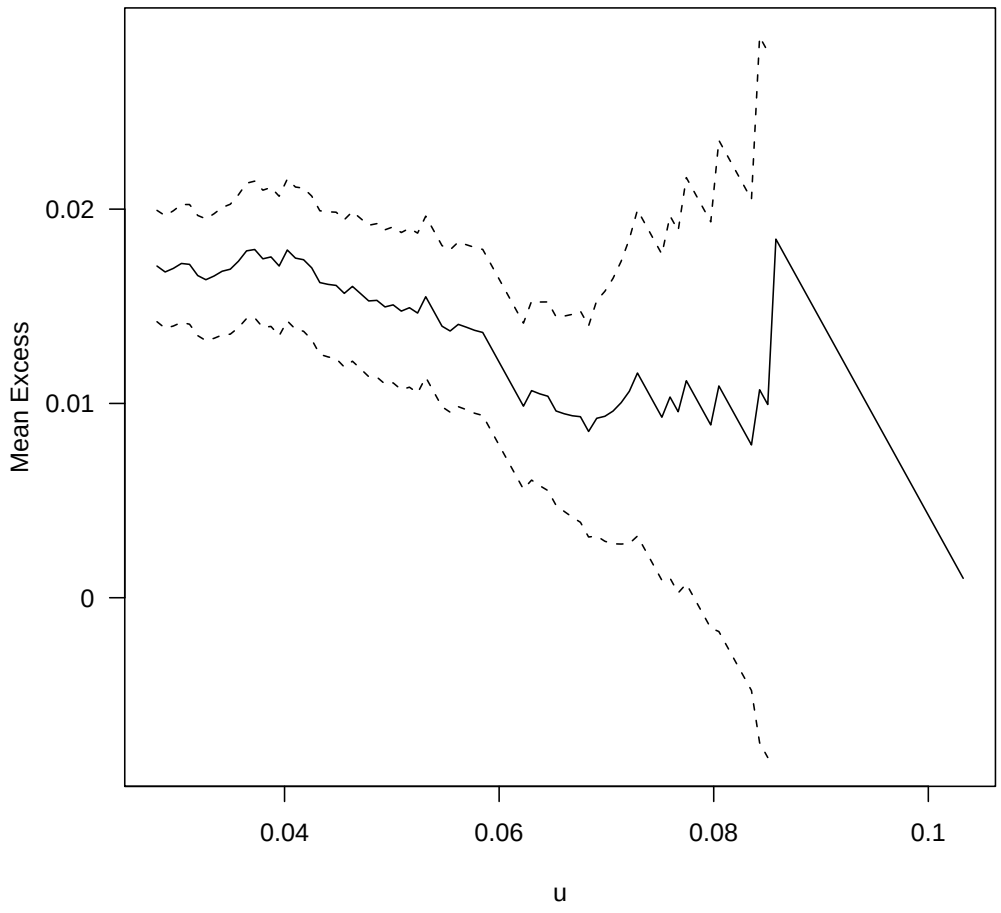

Figure 3.8: Mean residual life plot of estimated *sv*

The procedure to get a upper quantile estimate is as:

1. Define clusters of exceedences. The clusters of exceedences are identified as: The

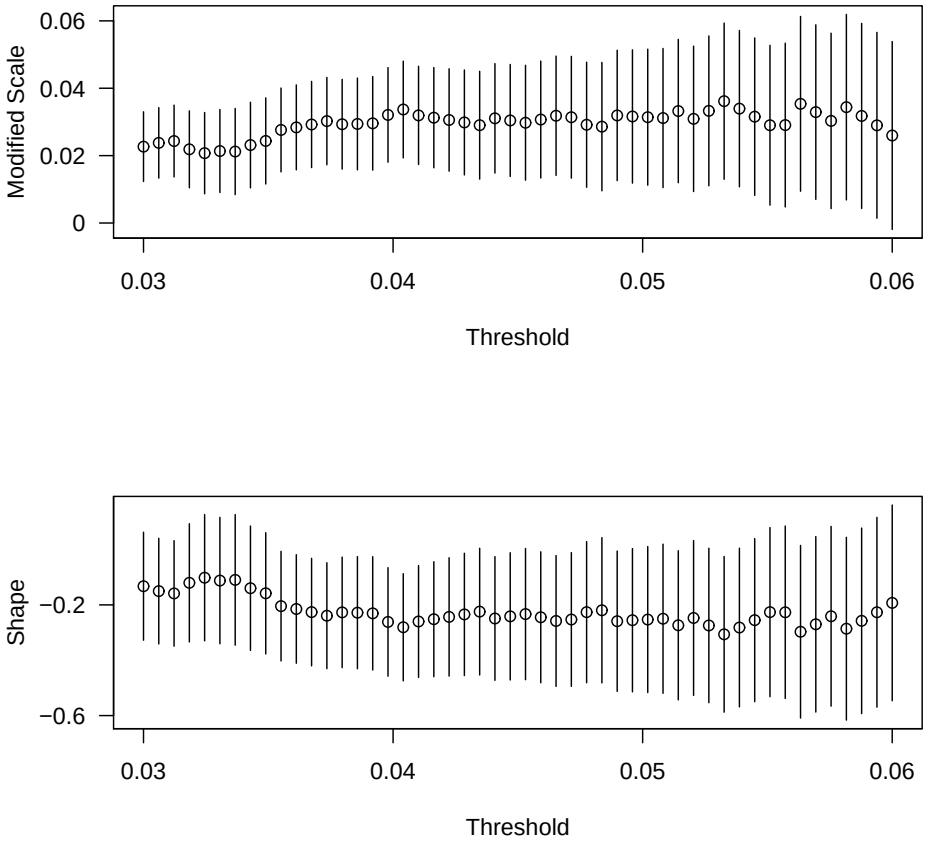

Figure 3.9: Parameter estimates against threshold for estimated sv

first exceedence of the threshold initiates the first cluster; The first cluster then remains active until either r consecutive values fall below (or are equal to) the threshold; The next exceedence of the threshold (if it exists) then initiates the second cluster, and so on. For $\{sv_t\}$ we chose $u = 0.038$.

2. Identify the maximum excess within each cluster and assume that cluster maxima are independent, with conditional excess distribution given by GPD.

3. Fit the GPD to the clusters maxima.

4. Estimate extremal index θ according to the method by Ferro and Segers (2003).

5. The upper quantile of estimated volatility at q, sv_q with $p(sv < sv_q) = q$, as:

$$sv_q = u + \frac{\sigma}{\xi}[(\frac{1}{1-q}\eta_u\theta)^{\xi} - 1]$$

where, $\eta_u = P(sv > u)$ is the probability of exceeding the threshold u and θ is the

estimated extremal index.

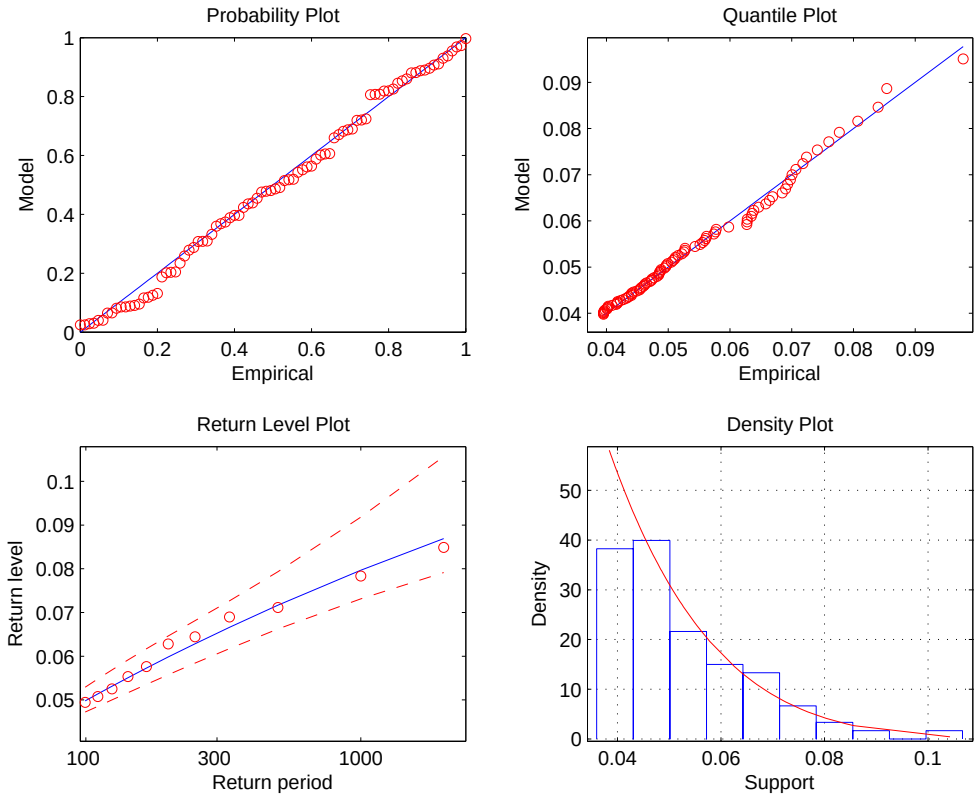

Figure 3.10: Diagnostic plot of GPD fitting for the upper tail of estimated sv

Figure 3.10 supply the diagnostic plots of the exceedence model fitting on $\{sv_t\}$. In the figure, the probability plot, quantile plot and density plot show a good fit of GPD to identified cluster maximums. The return level plot details the upper extreme quantile with the confidence interval of the quantile estimates. The solid line denotes the quantiles of estimated model and the small circles represent the empirical quantile values, which are very close to the model estimates. Figure 3.11 display the upper quantile estimates of the estimated stochastic volatilities $\{sv_t\}$ in using the method above. The high volatilities tend to appear consecutively and tend to increase once exceed a high threshold. The longer period which associated with large amount of consecutive high volatilities over a high quantile may indicate the potential instability resulted by the health status of preterm babies.

In the section we only supply a method in getting the estimates of the variability of preterm babies and finding the extreme quantiles of individual babies. Multiple samples from different babies and further statistical modelling of the random effects across multiple samples need to be studied in order to adequately describe the "abnormal" level of variability within the population. Other factors, such as the activity states, can affect the variabilities can be considered in further research. This study is only at the starting

stage of the neonatal research.

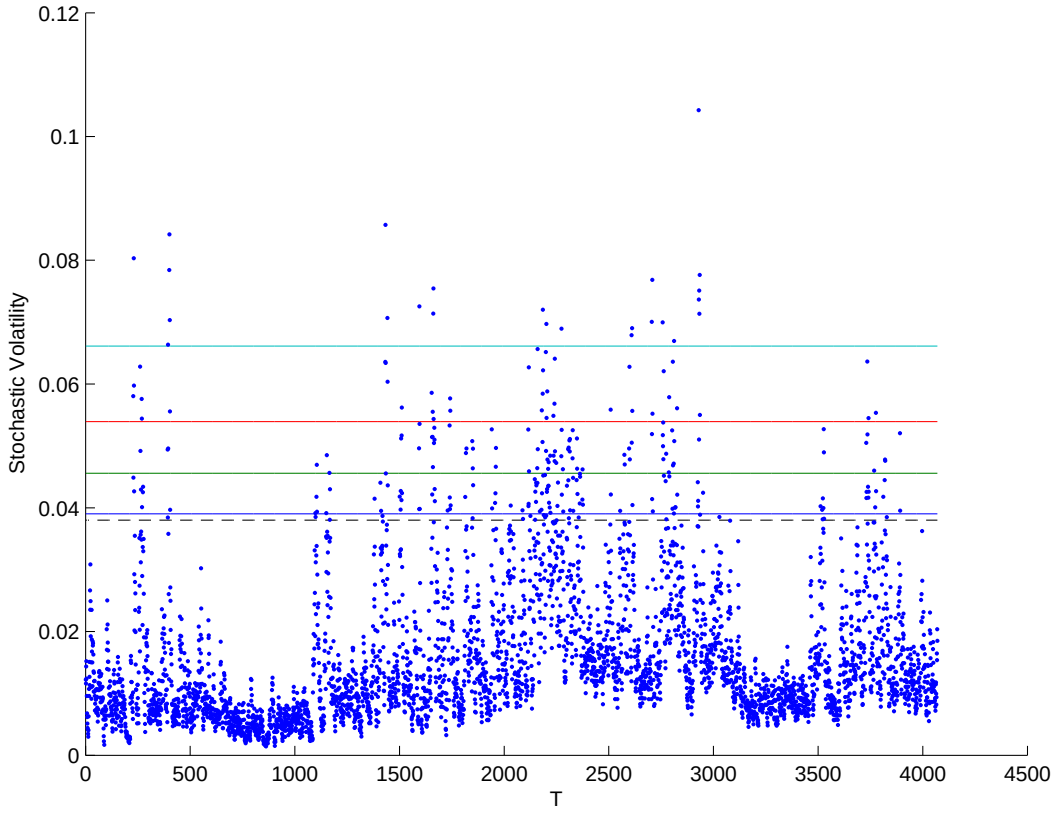

Figure 3.11: Upper quantiles of volatility. The solid horizontal line represent different quantiles as $sv_{98\%}, sv_{98.5\%}, sv_{99\%}, sv_{99.5\%}$. The dashed line denote the threshold value.

3.7 Summary

In this chapter, the stochastic volatility model implemented using an online sequential estimation, namely using an ASIR particle filter, is used to model the conditional volatility of oxygen concentration of preterm infants in order to help physicians determine the health status of preterm babies using the raw physiological measurements. Alternative volatility esitimators are considered to evaluate the performance of the SVM estimates, the results of which suggest that volatility estimation using SVM with particle filter is consistent with observed volatility. These empirical results show the stochastic volatility model provides a good estimator of the variability of the oxygen concentration data and therefore may be used to estimate the instantaneous latent volatility for the physiological measurements of preterm infants. Additionally, we apply extreme modelling with generalized pareto distribution to the upper tail of estimated stochastic volatilities to supply the high quantiles of the variability, which can be useful in identifying the potential risk of instability of preterm infants to help doctors in determine the healthy

statues of preterm babies. The dependence in the volatility estimates is dealt via POT method and extremal index used in the quantile estimation. Future research will need to consider how information across multiple patients can be pooled to provide physicians with estimates of normal ranges of the volatility at difference gestational within the population. This study provides insight into the stochastic volatility modeling approach and how it can be used to estimate the high quantiles (which are of interest for determining normal ranges) for a single patient.

Chapter 4

GARCH Dependence in Extreme Value Modelling

4.1 Introduction

Financial data is well known to exhibit relative heavy (heavier than normal) tails and extreme value theory has been shown to be a very useful tool in estimating and predicting the extremal behaviour of actuarial and financial products, such as predicting the largest claim in insurance and the Value at Risk (VaR), see Embrechts, Klüppelberg, and Mikosch (1997). However, applying extreme value models is not always straightforward because of the dependence behaviour often observed. For example, financial returns typically show clusters of observations in the tails, often termed volatility clustering.

There are three common approaches in accounting for the impact of the dependence in inferences. One of these approaches is to decluster the dependent extremes and then apply standard extreme models to the non-dependent sequence using statistical declustering algorithms, see Ferro and Segers (2003) for recent developments. Peaks-over-threshold (POT) is one such declustering approach (Davison and Smith 1990). Although the extremes still falls into the domain of attractions of GEV family, the extremal index need to be estimated for inference purpose. A second common approach is to apply standard extreme value models, but adjust the uncertainty estimates for the dependence using techniques like block bootstrapping, see Buishand (1993). However, the substantial drawback with both these approaches is although they do account for the dependence in the inferences, they do not provide additional understanding of the dependence structure.

Another common (two stage) approach in the finance literature (McNeil and Frey 2000;

Chan, Deng, Peng, and Xia 2007; Zhao, Scarrott, Oxley, and Reale 2010a), is to capture the dependence in the returns induced by the volatility clustering using a GARCH type model, followed by extreme value modeling of the approximately independent residual innovations. This two stage method overcomes the dependence in the extremes and explicitly captures the dependence structure itself, allowing further insight to be gained. The principal drawback with this approach is that accounting for the uncertainties associated with both stages is not completely straightforward.

In this paper, we explicit model the dependence behaviour of extremes via time varying extreme value parameters for a stationary extreme time series. The generalized auto-regressive conditional heteroscedastic (GARCH) process is widely used to model the volatility clustering observed in financial time series. We have amalgamated the GARCH model with the classical GEV model. Specifically, we consider a conditional GEV distribution with the specification that the extreme value sequence (maxima or minima) following an autoregressive lag one process with a GARCH type conditional variance structure. The dependence is captured by allowing the location and scale parameters of the GEV to follow a conditional autoregressive heteroscedastic process. With the combination of GEV and GARCH models, we will show that the combined model is better suited to explain the extreme quantiles than the classic GARCH model alone, which cannot capture the tail behaviours adequately with either normally distributed or even fatter tailed distributed (e.g. t) innovations as suggested by McNeil and Frey (2000). Also, an important benefit of our approach is that it is a one stage model and thus making inferences and accounting for all uncertainties straightforward.

Section 4.2 gives the review of classical GEV model for stationary sequences. The proposed GEV-GARCH model is outlined in Section 4.3. The estimation method of proposed model is discussed in Section 4.4. This study examines the dynamics of the maximum and minimum value of intra-day financial returns with focus on explaining the extremal quantiles, which can be useful in the control of risk. The study has been conducted with both simulated and real data, the results of which are presented in Sections 4.5 and 4.6. Further results on an extension to our approach considered by Bali and Weinbaum (2007) are provided in Section 4.7 considers what happens when the shape parameter of the GEV is also permitted to vary according to a GARCH type structure. It is demonstrated that model identification is very challenging if the GEV shape parameter is allowed to vary over time in such a manner, as one might expect due to the inherent difficulty in estimating this parameters due to the lack of information in the tails. Conclusions follow in Section 4.8.

4.2 Generalized Extreme Value Distribution

The generalized extreme value distribution (GEV) is an asymptotically motivated model for describing the distribution of the maxima and minima of a realized sequence of independent random variables. The GEV distribution has three parameters: the shape ξ, scale σ and location μ with distribution function:

$$G_{ev}(x|\xi,\sigma,\mu) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \quad \text{where } 1+\xi(x-\mu)/\sigma > 0 \qquad (4.1)$$

density:

$$g(x|\xi,\sigma,\mu) = \frac{1}{\sigma}G_{ev}(x|\xi,\sigma,\mu)\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-(1+1/\xi)} \quad \text{where } 1+\xi(x-\mu)/\sigma > 0 \quad (4.2)$$

The GEV model can represent three types of tail behaviour determined by the value of shape parameter ξ: Fréchet type (slowly decaying) tail ($\xi > 0$), the Weibull type (upper bounded) tail ($\xi < 0$) and Gumbel type (exponentially decay, like upper tail of normal distribution) which is defined in the limit as the shape parameter $\xi \to 0$. Extreme value theory shows that, if a limiting distribution for the maxima/minima of a sequence of independent and identically distributed random variables exists, then it must fall into one of these three types.

4.3 GEV-GARCH Model

The proposed model assumes that the observations come from the GEV distribution with a AR(1)-GARCH(1,1) process describing the conditional variance of extremes. Therefore, the distribution function of the observation x_t, representing the block maxima/minima, can be written as:

$$x_t \sim GEV(\xi,\sigma_t,\mu_t) \qquad (4.3)$$

$$G(x_t|\xi,\sigma_t,\mu_t) = \exp\left\{-\left[1+\xi\left(\frac{x_t-\mu_t}{\sigma_t}\right)\right]^{-1/\xi}\right\},$$

where $\sigma_t > 0$ and $1 + \xi(x_t - \mu_t)/\sigma_t > 0$, with the conditional parameter dependence:

$$\begin{aligned} x_t &= \beta_0 + \beta_1 x_{t-1} + \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \alpha_2 \epsilon_{t-1}^2 \\ \epsilon_{t-1} &= x_{t-1} - \mu_{t-1} \end{aligned}$$

where $|\beta_1| < 1$,$\alpha_0 > 0$,$\alpha_1 > 0$,$\alpha_2 > 0$, and $\alpha_1 + \alpha_2 < 1$ to ensure stationarity. Due to that the shape parameter requiring very large amounts of the data to get reliable estimates and the inherent lack of observation of extreme value modelling, it is generally unrealistic to try to model ξ as a function of time. Further, there is no empirical evidence shows that the tail behaviour changes over time. Therefore, the shape parameter ξ, as supported by empirical evidence (discussed in Section 4.7), is kept constant through time. A simulation study in Section 4.7 will show the model identification and estimation problems associated with a time varying shape parameter, as was proposed in related work by Bali and Weinbaum (2007).

The expected value and variance of x_t are:

$$E(x_t) = \mu_t + \sigma_t(\tau(1-\xi) - 1)/\xi, \quad \xi < 1 \tag{4.4}$$

$$Var(x_t) = \sigma_t^2(\tau(1-2\xi) - \tau^2(1-\xi)), \quad \xi < 0.5, \tag{4.5}$$

where outside of the ranges for the shape ξ the moments are undefined. The calculated $Var(x_t)$ is the conditional variance of extremes as it is conditional on the information up to $t-1$, which would be equivalent to an estimate of the volatility of the extremes.

4.4 Bayesian Inference

Bayesian inference is used for fitting the GEV-GARCH model as we can potentially take advantage of any expert prior information, which can be important in tail estimation due to the inherent sparsity of extremal data, and to account for all uncertainties in the estimation. Although, in this study we have deliberately used diffuse priors, to show our lack or prior information allowing the data to speak for themselves which demonstrates the worst case for estimation performance. Markov chain Monte Carlo (MCMC) has been used to obtain posterior distributions. Also the posterior distribution of Bayesian contain more information of the model parameter compare to other estimation methods since Bayesian analysis treat the parameter as a random variable and consider the uncertainty of the parameter in the inference.

4.4.1 Prior Distribution

The parameter vector $\theta = (\xi, \alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1)$ can be decomposed into two components $\theta_a = (\xi, \beta_0)$ which are defined over the whole real line and $\theta_b = (\alpha_0, \alpha_1, \alpha_2, \beta_1)$ which are bounded. We use a normal prior on θ_a and a flat prior on θ_b, to indicate little prior information, giving:

$$\pi(\theta_a) \propto \exp\left\{-\frac{1}{2}\left(\frac{\theta_a - \mu_{\theta_a}}{\sigma_{\theta_a}}\right)^2\right\}$$

$$\pi(\theta_b) \propto I_{(\theta_b)}$$

where, μ_{θ_a} and σ_{θ_a} are the location and standard deviation. $I_{(\theta_b)}$ is an indicator function which will reflect the constraints on the parameters defined in Section 4.3.

4.4.2 Posterior Distribution

Given $x_t \sim GEV(\xi, \sigma_t, \mu_t)$ and the density of GEV given in equation (4.2), The posterior distribution is:

$$p(\theta|x) \propto \prod_{t=1}^{T} g(x_t|\xi, \sigma_t, \mu_t) \times \exp\left\{-\frac{1}{2}\left(\frac{\theta_1 - \mu_{\theta_a}}{\sigma_{\theta_a}}\right)^2\right\} \times I_{(\theta_b)} \tag{4.6}$$

4.4.3 MCMC Procedure

A random walk Metropolis-Hastings algorithm has been used to sample the posteriors, as it has the advantage of being free of functional form since the posterior distribution function is not a proper probability function. MCMC has been used to update component by component, in order of importance of the parameters.

The M-H algorithm iterates the following steps in the study:

1. Initiate the parameter $\theta = (\xi, \alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1)$ values from the parameter space. Multiple starting value under the support of parameter values have been used in testing the convergency of the chain and the sensitivity to the initial value (Gelman, Carlin, Stern, and Rubin 2004). The simulation and empirical results both show the chain can converge (in probability) to the true value quickly even with the initial value which is relatively far from the true parameter value.

2. At iteration t, three steps are need:

   - *Draw a realization θ^* component by component in terms of the importance of the parameter, from the proposal distributions $q(\theta|\theta_{t-1})$. Since the shape*

parameter is the most important it is the first one to be updated. The components are therefore updated as follows: $\xi, \alpha_0, \alpha_1, \alpha_2, \beta_0$ *and* β_1.

- *The parameter restrictions within* θ_b *are enforced during the sampling procedure by rejecting the draws that violate them. The stationarity constraints,* $|\beta_1| < 1, \alpha_1 + \alpha 2 < 1$ *and constraint on range of support of the GEV* $1 + \xi(x_t - \mu)/\sigma_t > 0$ *is also imposed. Without those constraints the chain could simply converge to incorrect values or be computationally inefficient.*
- *Compute the acceptance probability, given as:*

$$\alpha(\theta^*|\theta_{t-1}) = \min\left\{1, \frac{p(\theta^*)/q(\theta^*|\theta_{t-1})}{p(\theta_{t-1})/q(\theta_{t-1}|\theta^*)}\right\} \tag{4.7}$$

Draw u *from the uniform(0,1), and set* $\theta_t = \theta^*$ *with* $u \leq \alpha$ *or* $\theta_t = \theta_{t-1}$ *with* $u > \alpha$.

3. Go back to step 2. Algorithm is iterated of step 2 a large number of times. The second half of the chain is used as the posterior distribution and the estimated parameter is calculated as the mean of highest posterior density interval (HPD) of the posteriors.

The converged chain is regarded as an approximate sample from the posterior distributions. The convergence of the final chain is checked by monitoring the marginal distributions of the parameters obtained from the parallel chains, using standard diagnostic checks, see Gelman, Carlin, Stern, and Rubin (2004).

4.5 Model Performance on Simulated Data

4.5.1 Parameter Estimations on Simulations

We first check the stability of the parameter estimation method using simulated data. Three different parameter sets were considered, with 100 simulated series for each parameter set with a sample size of 2500, which is of similar length to the real application sample in the following section. The first parameter vector set considered is close to those found for the application in Section 4.6. The second parameter set use a high coefficient of the lag variance term, α_1, and the third parameter use a relative lower value compared to the first set. The different shape parameter values used for these three sets are to identify the model features for heavier tails. Only positive shape parameters are considered as negative shape parameters are typically not observed in financial data.

An example sequence and associated distribution of the three parameter sets simulated samples are shown as Figure 4.1 and Figure 4.2. The sample parameter values are shown in Table 4.1 and sample 3 has the highest shape of GEV with a heavier upper tail than other two samples.

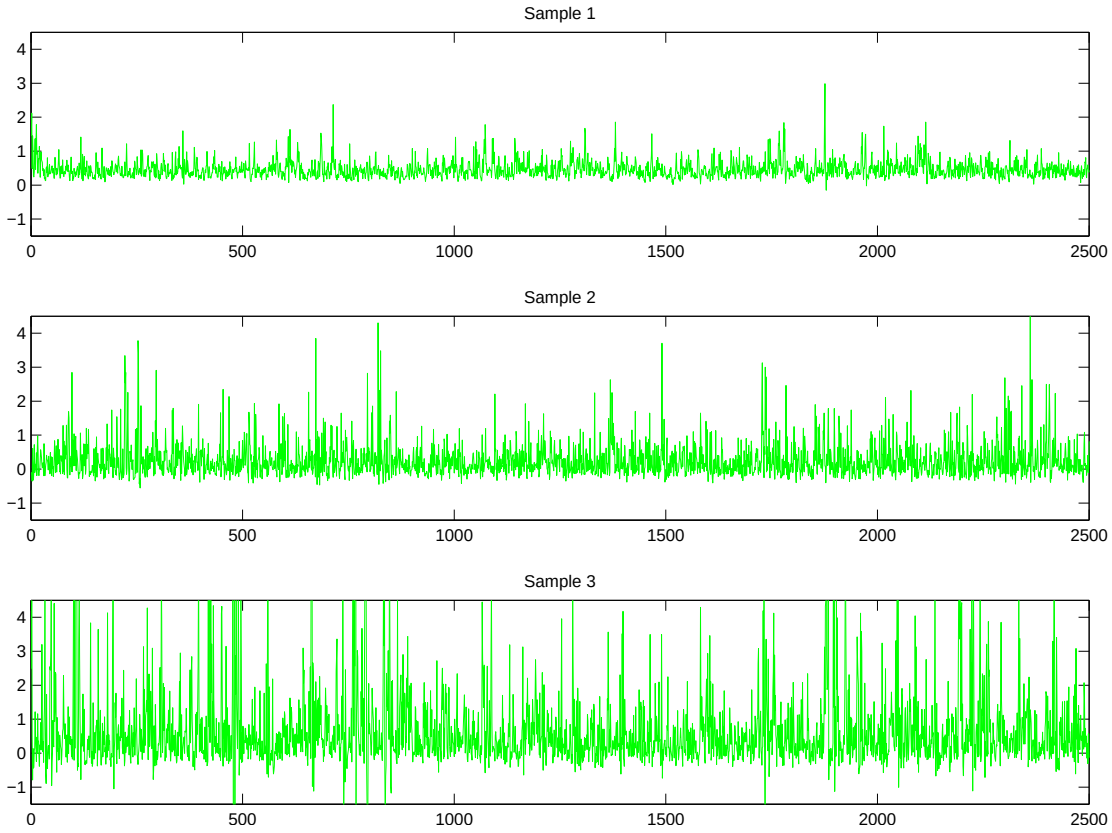

Figure 4.1: Example sequence of simulated samples. The simulated sample 1 to 3 use the parameter θ_1 to θ_3 in Table 4.1.

As mentioned above, in this study we use an relative uninformative prior, to indicate our lack of prior knowledge of these parameters and let the posterior driven by data themselves which demonstrates the worst case for estimation performance. Specifically, the prior for the simulation sample is as: $\xi \sim N(0.1, 0.15)$, $\alpha_0 \sim U(0, 0.2)$, $\alpha_1 \sim U(0.30, 0.99)$, $\alpha_2 \sim U(0, 0.5)$, $\beta_0 \sim N(0, 0.3)$, $\beta_1 \sim U(0, 0.5)$. Figure 4.3 to Figure 4.5 describe the posterior distributions of these three sample groups. Only one sample result is shown here for each simulation group for brevity. The estimated parameter value (mean over the highest posterior density (HPD) interval) and true parameter are consistent for all groups and the true parameter value are all contained by the credible interval of posteriors (95% HPD of posteriors). The difference between true parameter values and estimated values are due to nature sample variation.

Table 4.1 summarizes the parameter estimates results across all 100 simulations. The expected posterior predictive estimates are very close to the true values (indicating low

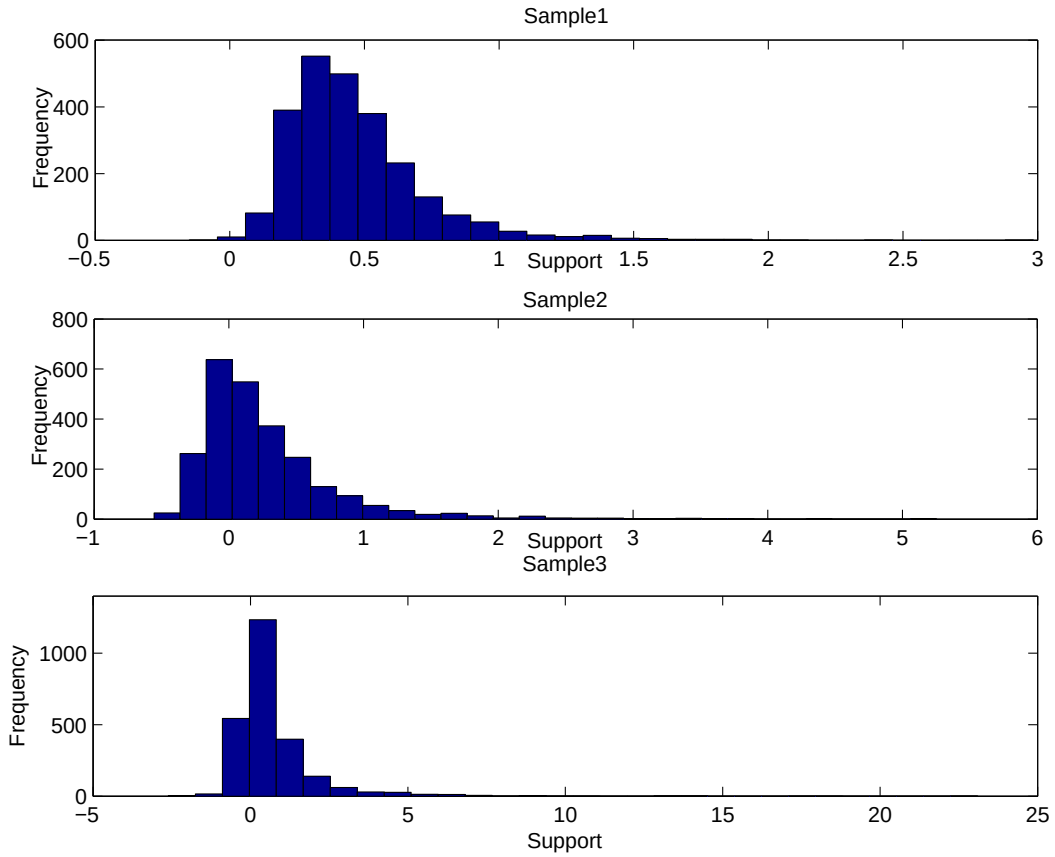

Figure 4.2: Example distribution of simulated samples. The simulated sample 1 to 3 use the parameter θ_1 to θ_3 in Table 4.1.

bias) and the sample variability is small. Generally, the root mean square error (RMSE) for ξ and α_1 (the coefficient of lag 1 term on the scale parameter) are higher than others, as expected, since they are the most difficult parameters to estimate. It is interesting to note that as the shape parameter value increases, the RMSE falls.

The expected variance of extreme values has been calculated according to Equation 4.4 and the extreme volatility is defined as the square root of the expected variance. The true extreme volatility is therefore can be obtained according to Equation 4.4 with true parameter value. The comparison of estimated and true extreme volatility, using simulation dataset 3, is shown Figure 4.6 and the result shows volatility estimates are very close to true values (i.e. low bias). Only an example of estimated volatilities (and true values) is given here for brevity, similar results for the rest simulated samples for parameter set 2 and other two simulation sets using parameter set 1&3.

Further, the extreme volatility estimates from the GEV-GARCH are compared with the volatility estimates from GARCH (particular, AR(1)-GARCH(1,1)) using the same simulations (using parameter 3) above. In Figure 4.7, there is no significate difference between two models in volatility estimators, since they have the same dependence struc-

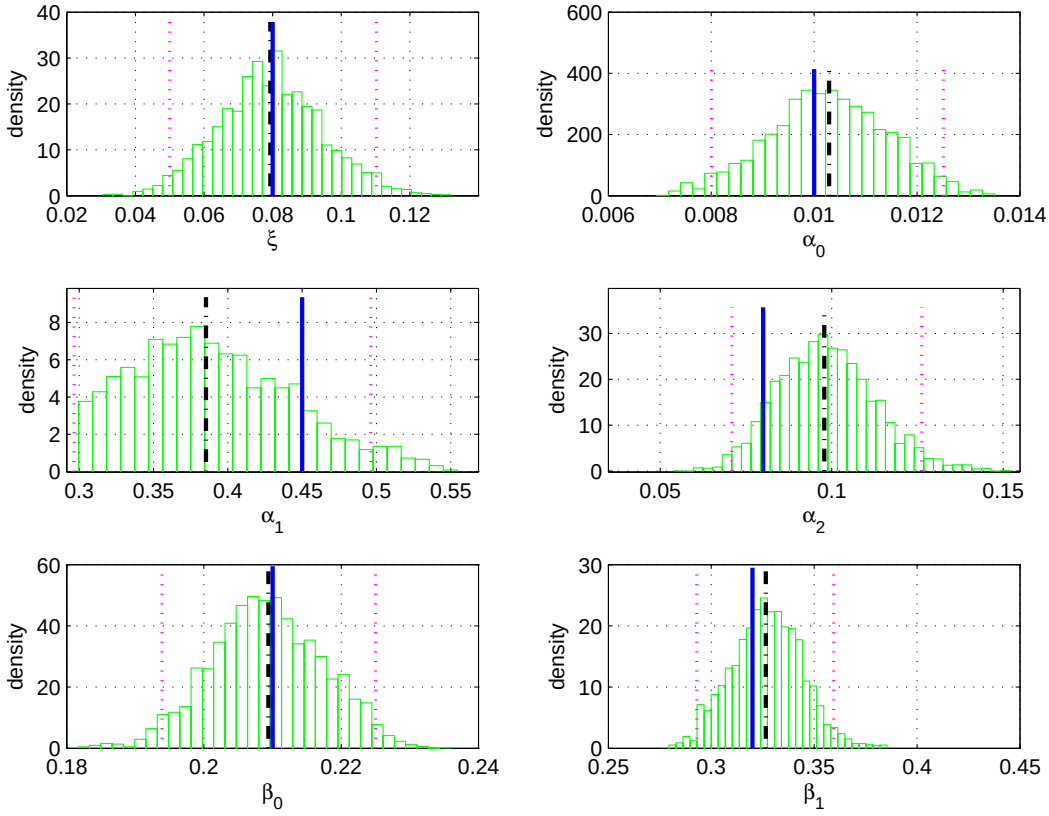

Figure 4.3: Posterior distribution of parameters for simulated sample with θ_1 in Table 4.1. The solid line indicate the true parameter values and the dashed line represent the mean of posteriors over the credible interval, which is denoted by dotted lines.

ture. The similarity in volatility clusters demonstrates the empirical consistency of the GARCH structure on the GEV parameters with the traditional GARCH variance estimates and suggests the dependence comes from the volatility clusters can be captured by proposed GEV-GARCH model.

4.5.2 Quantile Estimates of Simulated Data

The second part of our simulation study examine the quantile estimates of the model and compares the GEV-GARCH to other commonly used estimators.

The similarity of two volatility sequences in Figure 4.7 induces the question of what is the extra value of proposed GEV-GARCH model with respect to GARCH model alone? The GARCH is traditionally combined with a normal distribution for the innovations. The GEV-GARCH can better capture the extreme quantiles by combining the flexible extreme value tail behaviour and the conditional variance process. Figure 4.8 to Figure 4.10 plot the sequence of estimated quantiles using both the GEV-GARCH and GARCH (with assumed normally distributed innovations) models, along with the actual

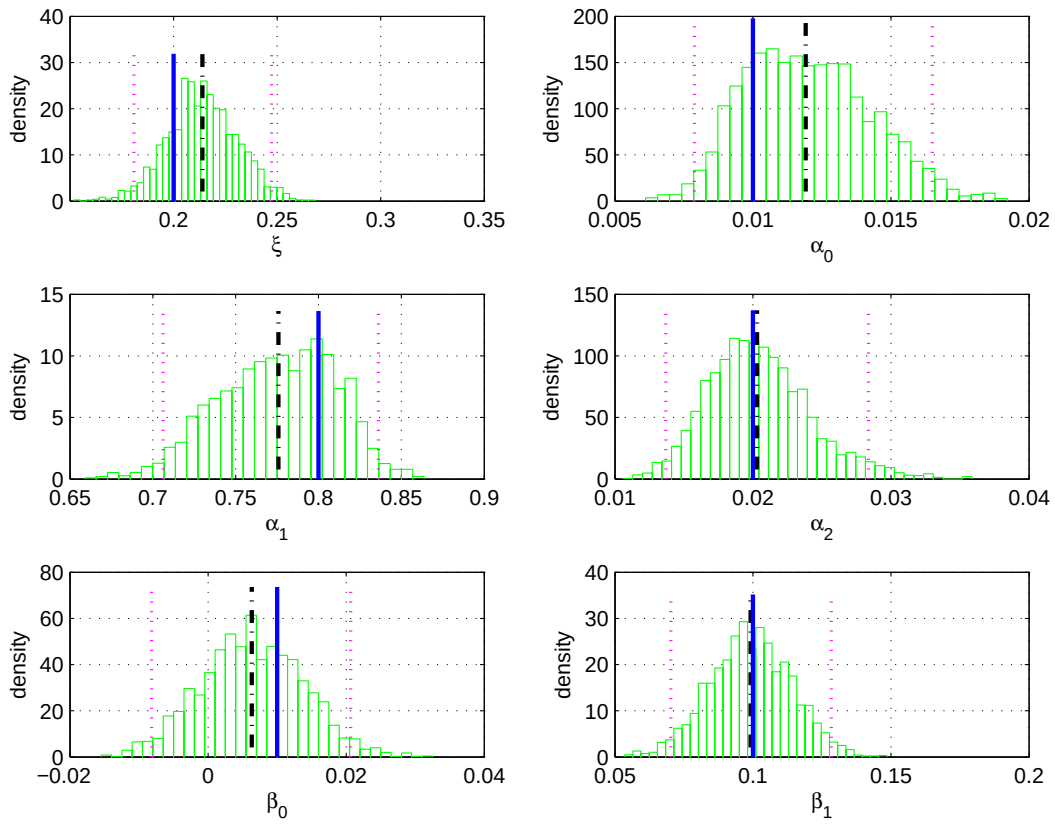

Figure 4.4: Posterior distribution of parameters for simulated sample with θ_2 in Table 4.1. The solid line indicate the true parameter values and the dashed line represent the mean of posteriors over the credible interval, which is denoted by dotted lines.

observations for this particular sample. Not all results are displayed for brevity. The extreme quantiles of the GEV-GARCH model are generally larger than the quantiles of GARCH model. Also as the further quantile is and the heavier the tail is, the larger difference of two quantile estimates is. It is clear that the extreme quantile estimates from the GEV-GARCH model has a better coverage rate of the extreme values than the corresponding GARCH quantiles, since the positive shape indicates a heavy tail from normal distribution typically assumed when using the GARCH.

We then compare the proposed GEV-GARCH model, to the two stage model approach of McNeil and Frey (2000) (with GPD for upper tail of innovations) and to a pure GARCH model (with normal innovations), in order to assess the general performance in terms of explaining the extreme quantiles. In the second stage of McNeil and Frey (2000) approach, they consider only the upper tail of the innovation distribution which can be effectively modelled using a generalized pareto distribution (GPD). See Section 2.5 for the quantile estimation in VaR with GPD. We will refer to the two stage quantile estimates using the two stage approach as the GARCH-GPD.

McNeil and Frey (2000) used maximum likelihood estimation when fitting the GPD to

Table 4.1: Results of fitting GEV-GARCH to 100 simulated datasets (with sample size of 2500). For each individual sample, the estimated parameter value is the mean of 95% highest posterior density interval (HPD) of the posterior samples. The estimated parameter value reported as the mean value over 100 samples and the RMSE is also provided.

θ_1	**TRUE**	**Estim.**	**RMSE**	θ_2	**TRUE**	**Estim.**	**RMSE**	θ_3	**TRUE**	**Estim.**	**RMSE**
ξ	0.08	0.072	0.024	ξ	0.2	0.201	0.016	ξ	0.3	0.2970	0.017
α_0	0.01	0.009	0.002	α_0	0.01	0.010	0.002	α_0	0.05	0.051	0.004
α_1	0.45	0.480	0.082	α_1	0.80	0.793	0.038	α_1	0.50	0.498	0.019
α_2	0.08	0.078	0.013	α_2	0.02	0.021	0.004	α_2	0.10	0.099	0.009
β_0	0.21	0.210	0.007	β_0	0.01	0.010	0.006	β_0	0.05	0.051	0.011
β_1	0.32	0.321	0.016	β_1	0.10	0.101	0.012	β_1	0.20	0.201	0.015

Table 4.2: RMSE of quantile estimates from 100 simulated datasets of length 2500, using the parameter sets defined in Table 1.

	Quantiles	**0.9**	**0.95**	**0.99**	**0.999**
θ_1	**GEV-GARCH**	0.0292	0.0399	0.0668	0.1117
	GARCH-GPD	0.0325	0.0459	0.0834	0.1542
	GARCH	0.0326	0.0852	0.2926	0.7280
θ_2	**GEV-GARCH**	0.0520	0.0760	0.1540	0.3630
	GARCH-GPD	0.1154	0.1639	0.3051	0.6110
	GARCH	0.1347	0.1862	0.7951	2.5168
θ_3	**GEV-GARCH**	0.0372	0.0560	0.1701	0.6891
	GARCH-GPD	0.3170	0.4416	1.1870	4.3459
	GARCH	0.2238	0.7266	3.4622	12.0951

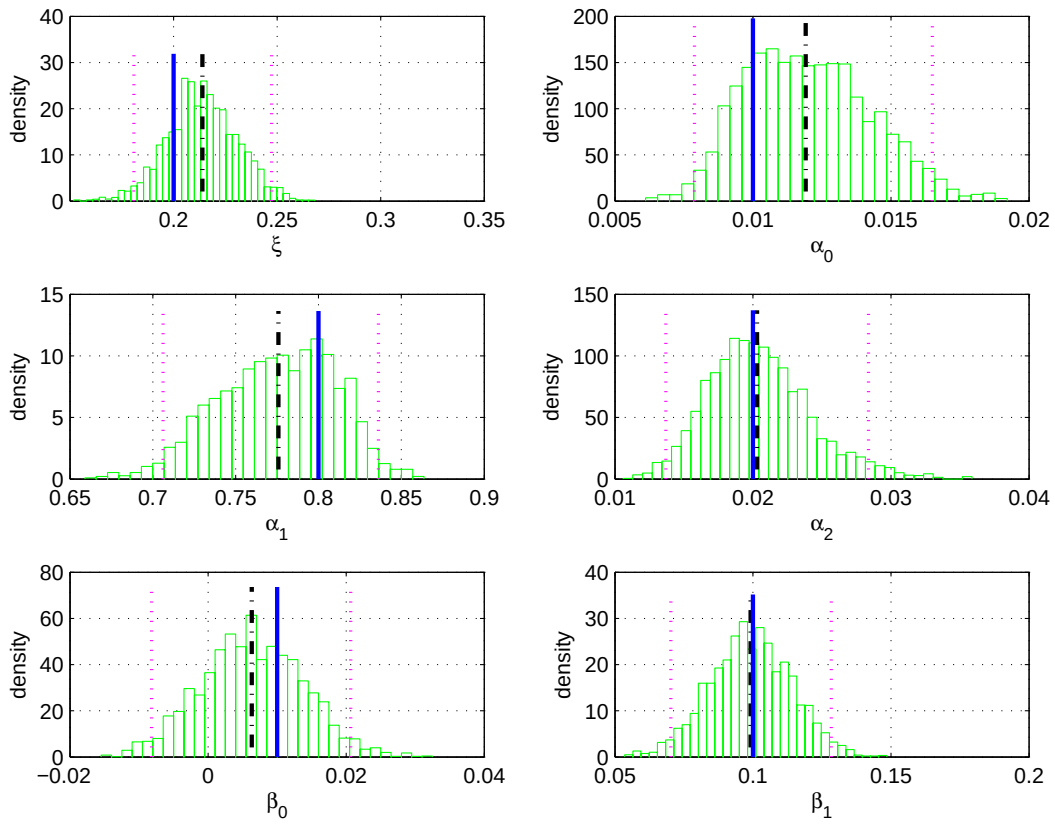

Figure 4.5: Posterior distribution of parameters for simulated sample with θ_3 in Table 4.1. The solid line indicate the true parameter values and the dashed line represent the mean of posteriors over the credible interval, which is denoted by dotted lines.

the upper tail, whereas we use Bayesian inference (except for η_u will be well estimated using the sample proportion) for direct comparison to the GEV-GARCH. The threshold choice can be problematic for GPD (Mendes and Lopes 2004; Tancredi, Anderson, and O'Hagan 2006; Zhao, Scarrott, Oxley, and Reale 2010a), however discussion of these issues is covered in detail in Chapter 5. We simply chose to use the upper 25% of the data. A sensitivity analysis was carried out to assess this pre-chosen threshold, which showed no spurious artefacts were introduced. We carried out a complete simulation study with 100 simulated datasets for each parameter set mentioned in Table 4.1. The key results are shown in Table 4.2, which compares RMSE of the quantile estimates based on the three different models.

In Table 4.2 the RMSE for all three models increases with the quantile as expected, due to the reduced information further into the tails. Note that the data is simulated from the GEV-GARCH model, so we expect this model should have the lowest RMSE, as observed in Table 4.2. Notice that the GEV-GARCH and GARCH-GPD are very similar in RMSE, and they both outperform the standard GARCH implementation. These results show that the commonly used two-stage approach of McNeil and Frey (2000) and our

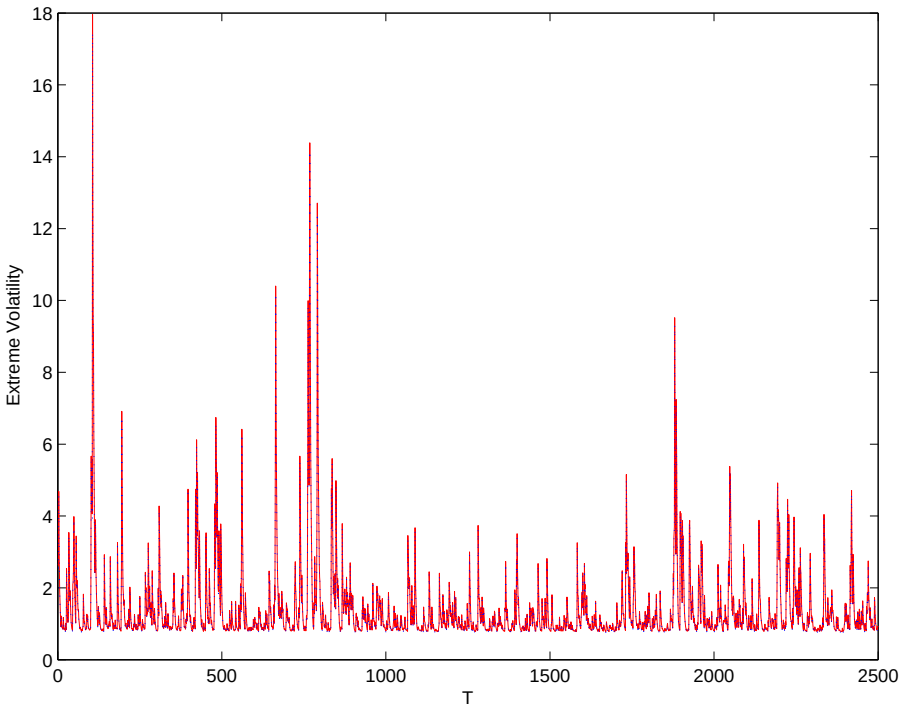

Figure 4.6: Estimated and true extreme volatilities of simulated dataset with θ_3 in Table 4.1. The dashed line represents the estimated volatilities from GEV-GARCH model and the solid line represents the true extreme volatilities.

proposed approach have similar performance for estimating high quantiles. However, our approach has an important benefit that it is a one stage model which simplifies the parameter estimation and makes it easier to account for all sources of uncertainty, which is much more challenging in the two stage approach. All the other simulated samples have similar results. These results suggest that the GEV-GARCH should be preferred if the interest is in modelling or predicting extremes with conditional variance.

4.6 Empirical Results

Bali and Weinbaum (2007) develop a similar conditional extreme value volatility estimator based on high frequency returns which allows all three parameters of GEV distribution to vary over time. They apply the model to daily 5 minute maximum returns of stock index to compare the relative performance of with GARCH and implied volatility (VIX) in forecasts of realized volatility. We will return to assess the performance of their approach in Section 4.7. In this study, the GEV-GARCH model is also applied to daily 5 minute maximum of individual stock returns. The data used in this section is kindly provided by Scharth and Medeiros (2006), including intra-day all transactions of 4 Dow Jones Industrial Average index stocks as: Alcoa, General Electric, IBM and Coca-Cola.

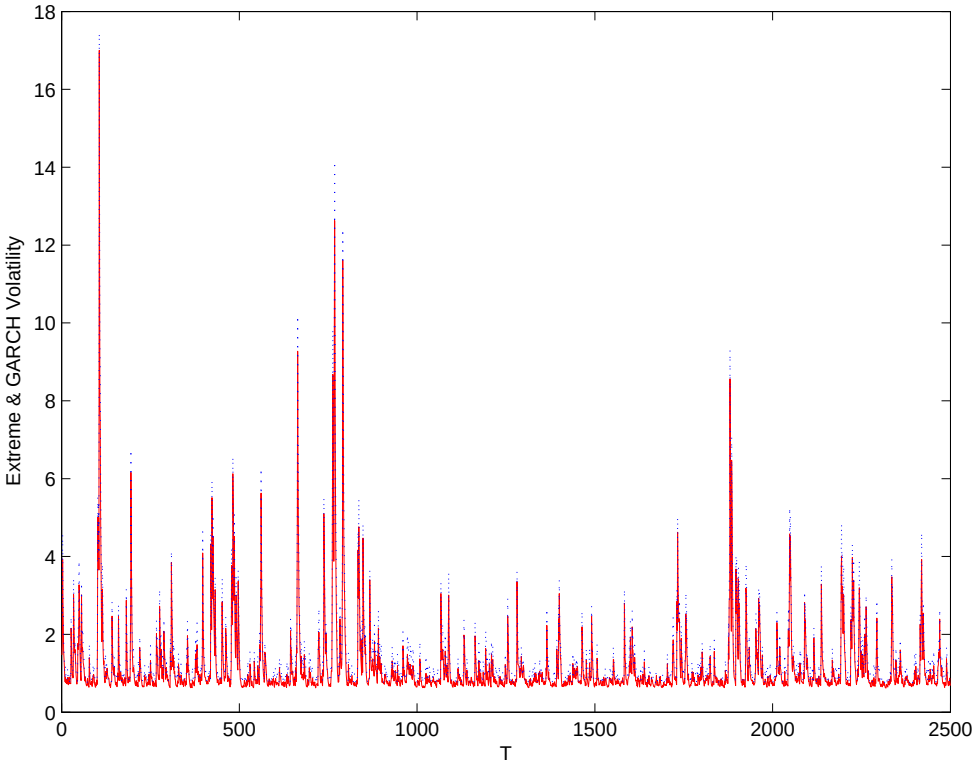

Figure 4.7: GARCH and GEV-GARCH extreme volatilities of simulated dataset with θ_3 in Table 4.1. The dashed line represents the volatilities from GEV-GARCH model and the solid line represents the volatilities from the GARCH (with normal innovations) model.

The period is from January 3, 1994 to December 31, 2003, leaving a total of 2540 days. We only show the results for IBM for the period collected in the section for brevity.

We start sampling the block maxima/minima by removing non-standard quotes, filtering possible errors and obtaining valid transactions for the 9:30 am to 16:05 p.m. period. Followed by Aït-Sahalia, Mykland, and Zhang (2005), we use two time scale to construct 5 minute log return to minimize the microstructure noise and the daily maxima/minima of them is used in the following analysis. We also tried sampling daily block maxima/minima using 2 minute and 10 minute return to find the impact of frequency on the distribution of tails. The 2 minute frequency will not induce a reliable extreme observations and 10 minute return results are similar as using 5 minute. Only the results associated with 5 minute frequency on IBM are reported in the following sections. The frequency cannot be too low (eg. longer than 30 minute), which results no enough daily observations to define extremes, hence losing the tail information. The choice of frequency should consider the application interest and representability of extremes.

Figure 4.11 shows the distribution of daily 5min return maxima of IBM and clearly there is a heavier upper tail. Figure 4.12 confirms a heavy-tail distribution of the daily

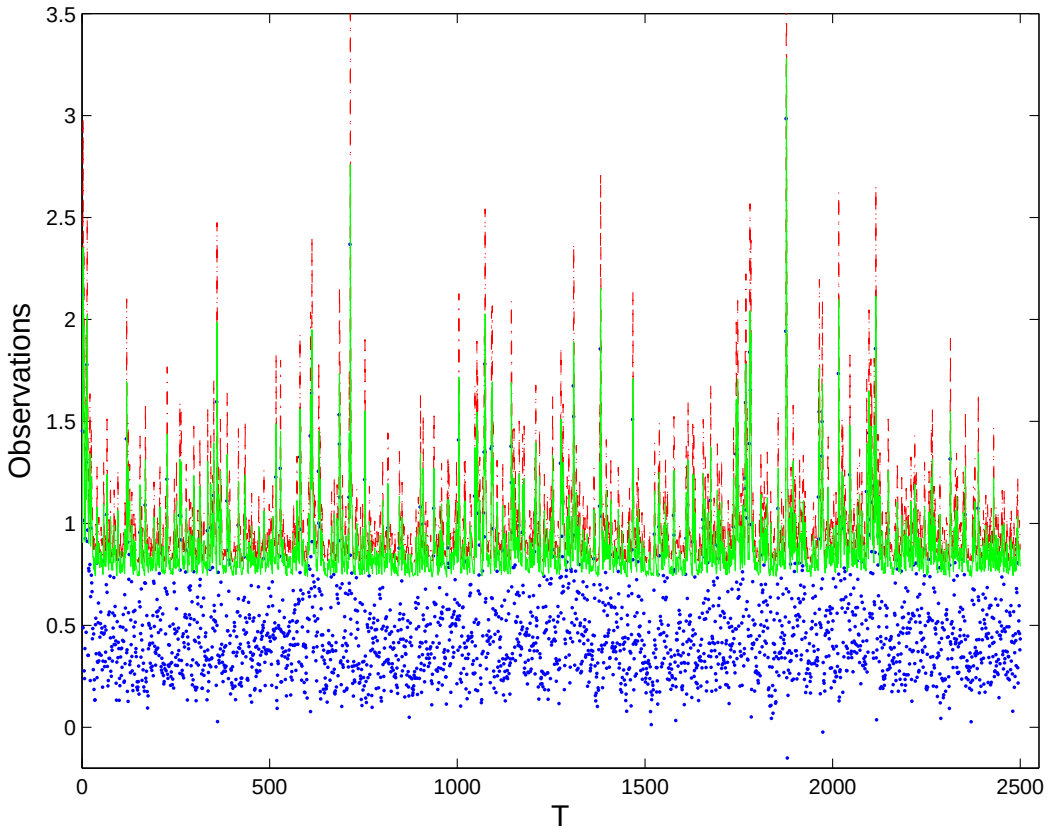

Figure 4.8: Simulated sample with θ_1 in Table 4.1 with 97% quantiles estimates from GEV-GARCH and GARCH (with normal innovations). The dots indicate the simulated data points. The dashed line represents the quantiles from GEV-GARCH model and the solid line represents the quantiles from GARCH model (with normal innovations).

maxima by comparing the empirical quantiles with normal quantiles. The estimated results for stock return are very similar to the simulation results of θ_1 reported in the previous section. Table 4.3 reports the parameter estimates for the IBM daily 5 minute maximum returns and 95% credible interval of the posterior obtained using MCMC. The shape parameter is positive as expected and the posterior distribution is shown in Figure 4.13.

The GARCH coefficient α_1 is less than 0.5. The time varying scale σ_t and location μ_t in time of GEV are plotted in Figure 4.14. Then the extreme volatility can be obtain with time varying σ_t as in Figure 4.15 along with the estimated GARCH volatilities. The similarity of two volatility estimators indicates that the GEV-GARCH model has captured the volatility clustering via imposing a time varying scale parameter.

When the quantile estimates are of concern as Figure 4.16, the GEV-GARCH expected quantile (of 97%) has a better coverage rate than the the GARCH (with normal innovations) due to its flexibility in capturing the heavy tail behaviour. Also we can expect a even larger difference between two quantile estimators when compare higher quantiles

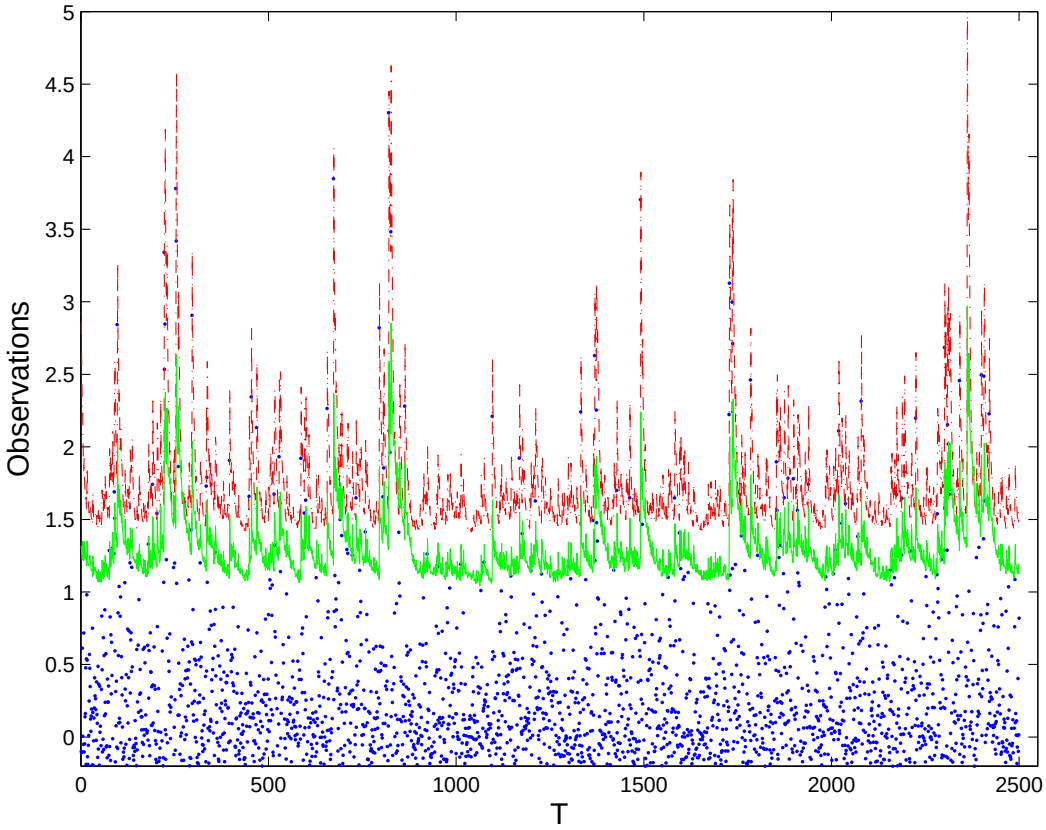

Figure 4.9: Simulated sample with θ_2 in Table 4.1 with 98% quantiles estimates from GEV-GARCH and GARCH (with normal innovations). The dots indicate the simulated data points. The dashed line represents the quantiles from GEV-GARCH model and the solid line represents the quantiles from GARCH model (with normal innovations).

since the tail quantile distributions are more skewed and heavy tailed at higher quantiles which is showing the asymmetry in the information (and lack of information) when estimating higher quantiles.

4.7 Exploration of Time-varying Shape Parameter

In the proposed GEV-GARCH model, the shape parameter of the GEV distribution is treated as constant over time. It is well known that the shape parameter is very difficult to estimate and requires very large datasets, see Coles (2001), to get reliable estimates. Further, it is unclear whether it is physically meaningful to allow the shape to vary over rapidly time, as it is driven by marketing trading behaviour. Bali and Weinbaum (2007) consider a related extension of the GEV to that proposed here, where they also allow the shape parameter to vary in time according to the structure:

$$x_t \sim G(\xi_t, \sigma_t, \mu_t)$$

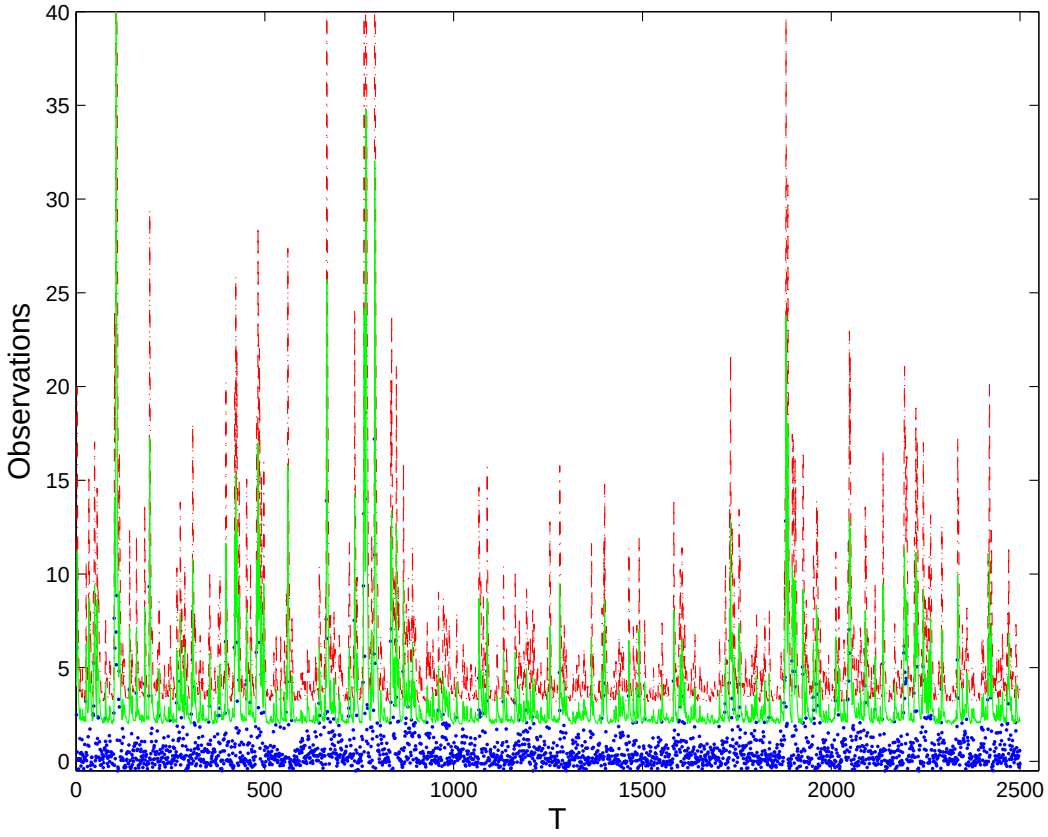

Figure 4.10: Simulated sample with θ_3 in Table 4.1 with 99% quantiles estimates from GEV-GARCH and GARCH (with normal innovations). The dots indicate the simulated data points. The dashed line represents the quantiles from GEV-GARCH model and the solid line represents the quantiles from GARCH model (with normal innovations).

$$\begin{aligned} \mu_t &= \mu_0 + \phi x_{t-1} \\ \sigma_t &= \sigma_0 + \lambda_1 \sigma_{t-1} + \lambda_2 |\epsilon_{t-1}| \\ \xi_t &= \xi_0 + \gamma_1 \xi_{t-1} + \gamma_2 |\epsilon_{t-1}| \end{aligned}$$

where, $\epsilon_{t-1} = x_{t-1} - \mu_{t-1}$, $1 + \xi(x_t - \mu_t)/\sigma_t > 0$, and $|\beta_1| < 1$,$\alpha_0 > 0$,$\alpha_1 > 0$,$\alpha_2 > 0$, and $\alpha_1 + \alpha_2 < 1$ to ensure stationarity.

In attempting to asses a time varying structure of ξ proposed by Bali and Weinbaum (2007) and compare it with our approach, we tried to carry out a simulation study based the their model (denoted as Bali's model following). We firstly tried simulate samples exactly as their model form and then simulate the data as adding a time varying shape of GEV on our proposed GEV-GARCH model (denoted as GEV-GARCH-ξ following) above as:

$$x_t \sim G(\xi_t, \sigma_t, \mu_t)$$

Table 4.3: Mean of posterior parameter estimates for 5 minute returns of IBM. *Note:* the CI is the 95% highest posterior credible interval of the posterior.

θ	**Estimated Value**	**CI**	
		Low	**High**
ξ	0.076	0.048	0.104
α_0	0.006	0.004	0.007
α_1	0.453	0.309	0.598
α_2	0.083	0.055	0.113
β_0	0.207	0.193	0.221
β_1	0.32	0.286	0.356

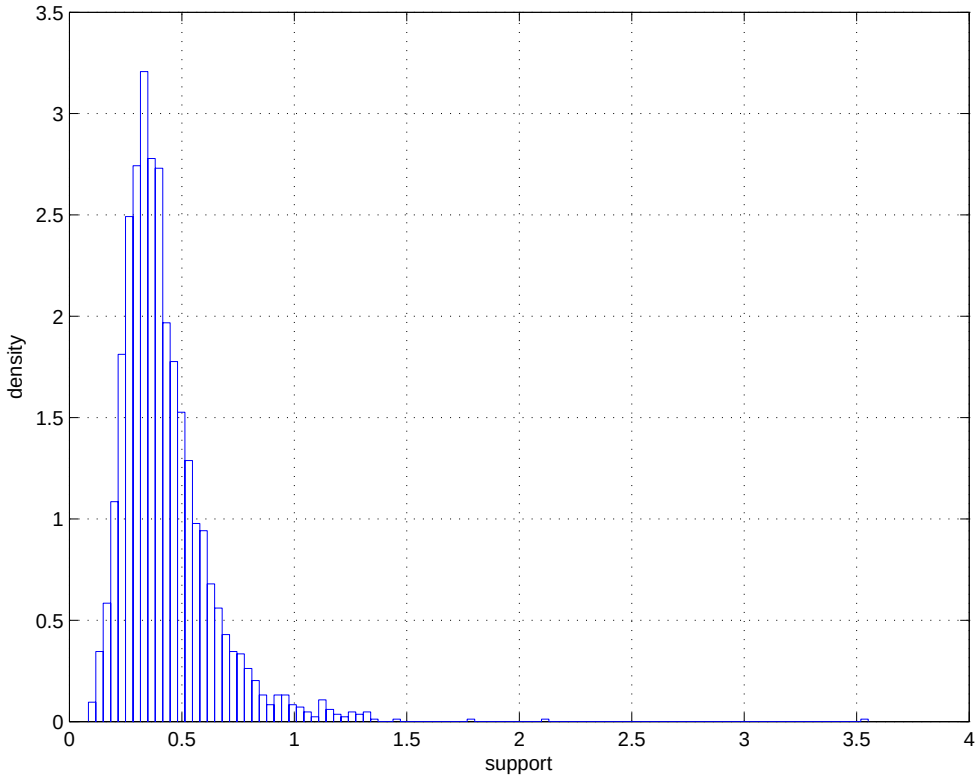

Figure 4.11: Empirical distribution of IBM daily maxima (5 minute maximum).

$$\begin{aligned} x_t &= \beta_0 + \beta_1 x_{t-1} + \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \alpha_2 \epsilon_{t-1}^2 \\ \xi_t &= \xi_0 + \gamma_1 \xi_{t-1} + \gamma_2 |\epsilon_{t-1}| \end{aligned}$$

where, $\epsilon_{t-1} = x_{t-1} - \mu_{t-1}$, $1 + \xi(x_t - \mu_t)/\sigma_t > 0$, and $|\beta_1| < 1$,$\alpha_0 > 0$,$\alpha_1 > 0$,$\alpha_2 > 0$, $\alpha_1 + \alpha_2 < 1$ to ensure stationarity.

For both of simulations, ML estimation frequently failed due to problems of identifiability of the shape parameter. To demonstrate, Figure 4.17 shows the profile likelihood for the shape related parameters γ_1 and γ_2 from a simulated dataset based on Bali's model

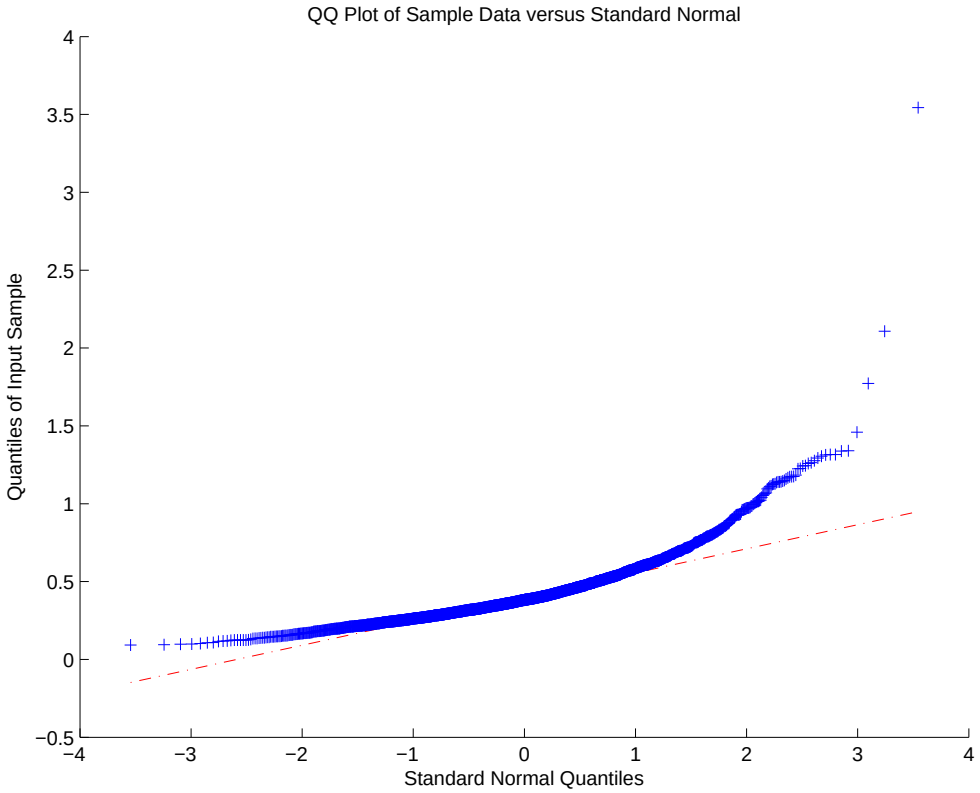

Figure 4.12: Empirical Quantile vs Normal Quantile of IBM daily maxima.

with parameter vector $(0.005, 0.87, 0.045, 0.001, 0.93, 0.032, 0.001, 0.28)$ (parameter $\theta = (\lambda_0, \lambda_1, \lambda_2, \xi_0, \gamma_1, \gamma_2, \mu_0, \phi)$) as the empirical results in Bali and Weinbaum (2007). You will notice that the profile likelihood for γ_1 is almost flat over the sensible axis range plotted for the simulated sample. Various simulated dataset with different parameter sets were considered and the likelihood are either flat or having problem to converge, even with a initial value very close to the true parameter values.

Figure 4.18 shows the profile likelihood for the shape related parameters γ_1 and γ_2 from a simulated dataset based on GEV-GARCH-ξ model with parameter vector $(0.01, 0.85, 0.05, 0.01, 0.8, 0.01, 0.01, 0.28)$ (parameter $\theta = (\alpha_0, \alpha_1, \alpha_2, \xi_0, \gamma_1, \gamma_2, \beta_0, \beta_1)$). Again a flat likelihood surface indicates the difficulty in identifing the parameters.

Hence, even for the correct model with a sample of size 3000 it appears that the likelihood estimation cannot identify the shape related parameters. Various simulated datasets with different parameter sets were considered and very similar results produced. The rest of results are not reported here for brevity.

We then apply both models with time varying shape of GEV on the IBM daily maxima sequence above. Figure 4.19 gives the profile likelihoods contour of Bali's model and Figure 4.20 supplies the likelihood contour of GEV-GARCH-ξ model. You will notice that the profile likelihood has multiple modes for both models, which would clearly

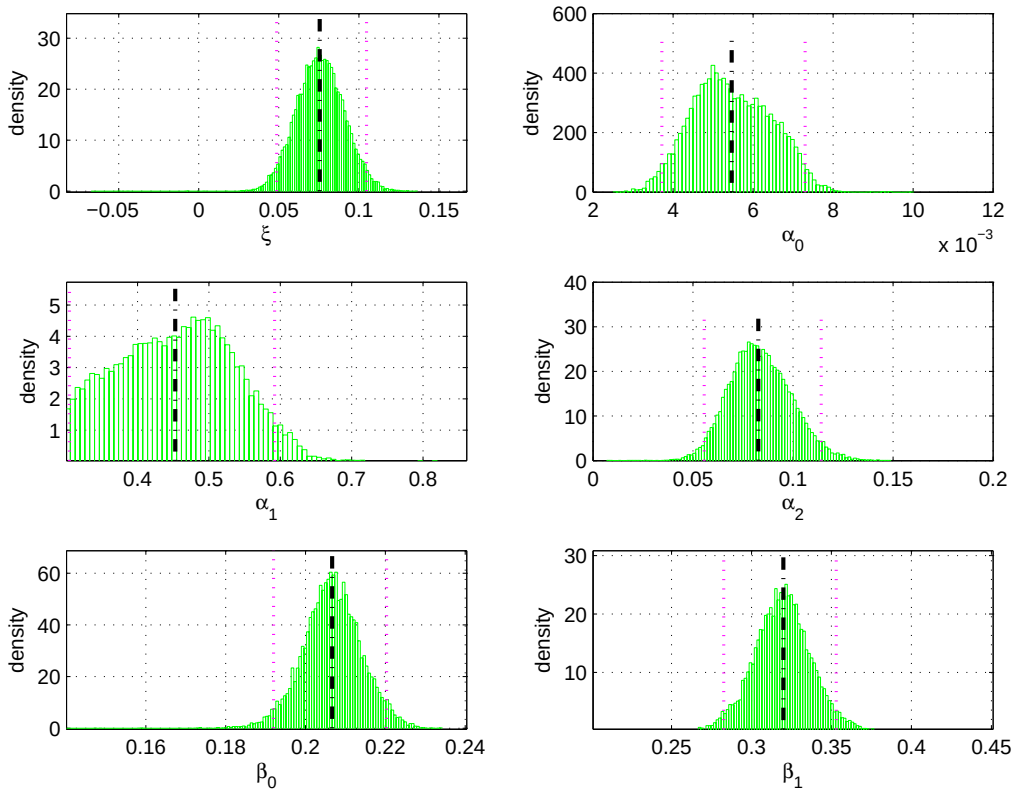

Figure 4.13: Posterior distribution of IBM daily maxima. The dotted two vertical lines are the 95% credible interval and the dashed vertical line denote the mean of posterior on the credible interval.

cause problems for standard numerical optimization procedures used to maximize the likelihood, and is indicative again of the problem of model identifiability when the shape parameter is permitted vary in time in such dynamic manner. Both simulation and empirical results on the model frame with a time varying shape of GEV suggest model identification problem.

The simulations and empirical results show the model identification problem induced by a time varying shape of GEV. Physically there is no empirical evidence that shape of distribution of extreme returns can change that fast as the conditional variances.

4.8 Summary

We have proposed a new extreme value model for dynamic estimation of extremal quantiles. To capture the temporal dependence typically observed in financial applications, that of volatility clustering, the scale and location parameters of the generalized extreme value distribution (GEV) have been defined by a conditional autoregressive heteroscedastic structure. Arguably, the most commonly used approach in the finance literature to

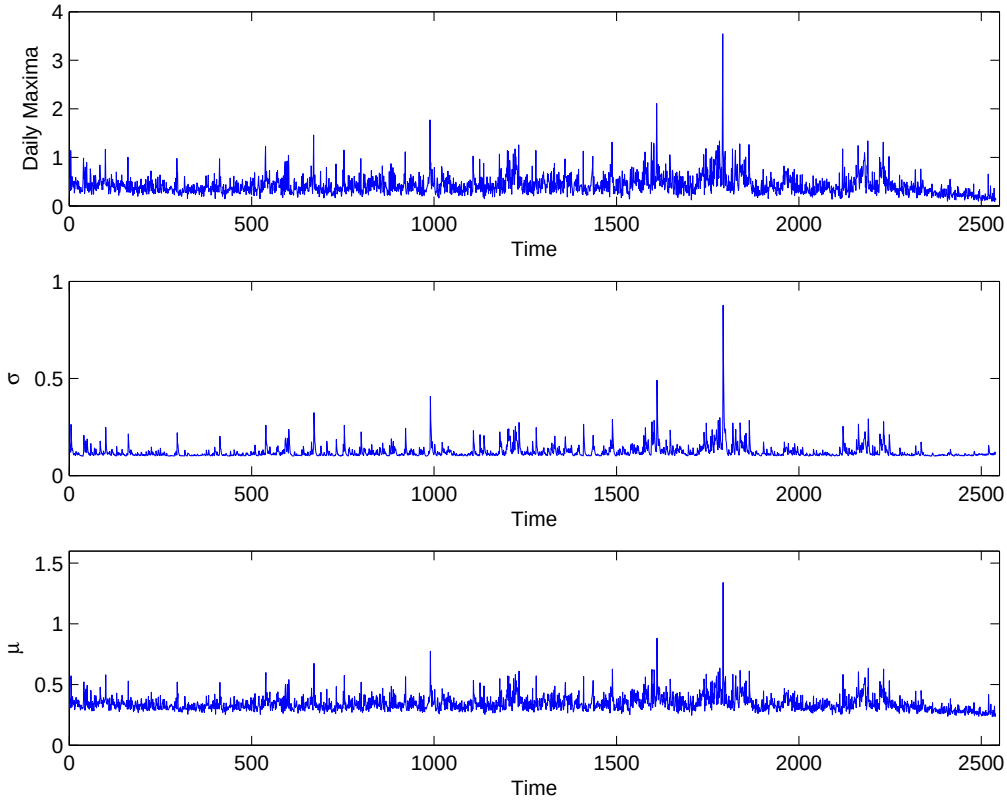

Figure 4.14: IBM GEV parameter time series with daily maxima series

overcome this temporal dependence is the two stage approach of McNeil and Frey (2000), where the dependence is captured by a GARCH model (or one of it's many variants) followed by applying standard extreme value modelling approach to the approximately independent residual innovations. The proposed approach overcomes the challenges associated with uncertainty estimation in this two stage approach, and also avoids the problems associated with threshold determination.

A simulation study and real data application have been used to demonstrate the performance of the GEV-GARCH in capturing the dynamic of conditional variance of extremes and to model the tail behaviour of the underlying variables. The GEV-GARCH performance in estimating the conditional variance is compared to the standard GARCH model with normal residual innovations, and is shown to provide estimates with much smaller RMSE. The performance in estimating the extremal quantiles is compared the two stage approach of McNeil and Frey (2000) and standard GARCH approach with normal residual innovations. The proposed model is shown to outperform the standard GARCH for estimating the tail quantiles, due to flexibility of the extreme value distribution for capturing the heavy tail behaviour typically observed in financial times series. The GEV-GARCH provides similar performance to the two stage approach where the generalized Pareto distribution is used to describe the tail of the innovation distribu-

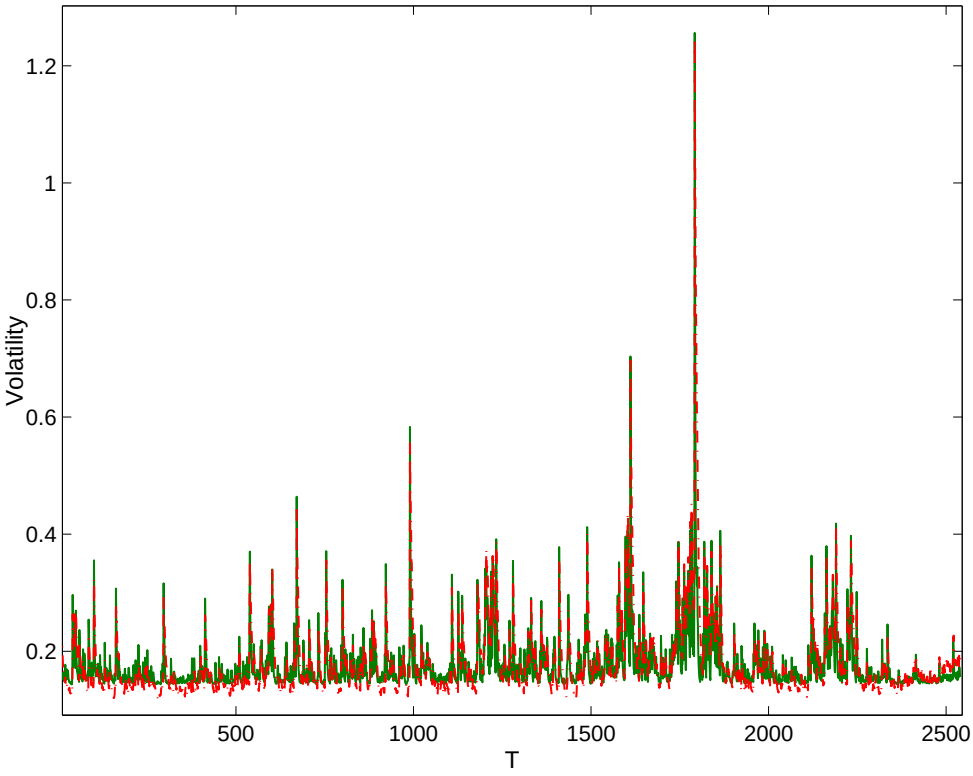

Figure 4.15: Volatilities of IBM return (5 minute maximum). The solid line represents the volatilities from GEV-GARCH model and the dashed line represents the volatilities from the GARCH model (with normal innovations).

tion, but has an important benefit in that it is a one stage procedure so making the parameters estimation and accounting for the uncertainties more straightforward.

A final simulation study and empirical study is used to demonstrate model identification and parameter estimation complications from permitting a time varying shape parameter with a similar GARCH structure, as proposed by Bali and Weinbaum (2007). The shape parameter is known to be challenging to estimate requiring very large sample sizes, see Coles (2001). It is possible that the real structure of shape could change over time, for example if the market structure changes. However, it is clear in the results presented that model identification is problematic if the tail behaviour changes as fast as the volatilities. Further work could investigate a smoothly varying or structural break type model (Davison and Ramesh 2000b; Conigliani and Tancredi 2005) for the shape parameter if there is believed to be some form of market change.

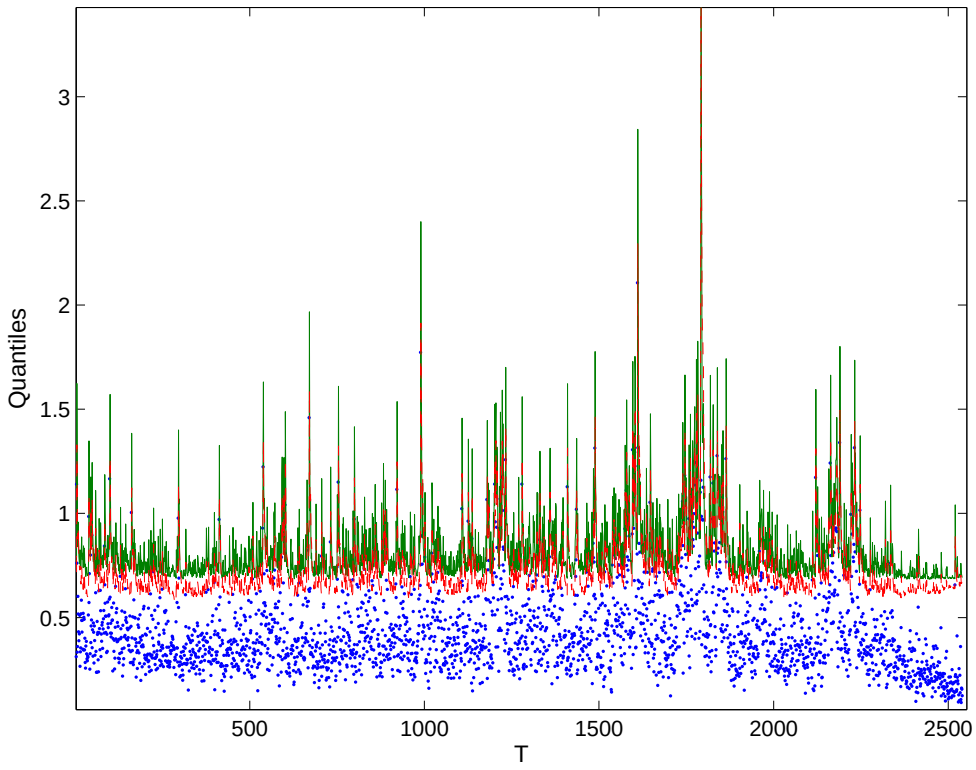

Figure 4.16: Estimated 97% quantiles of IBM returns (5 minute maximum). The solid line represents the quantiles from GEV-GARCH model and the dashed line represents the quantiles from the GARCH model (with normal innovations).

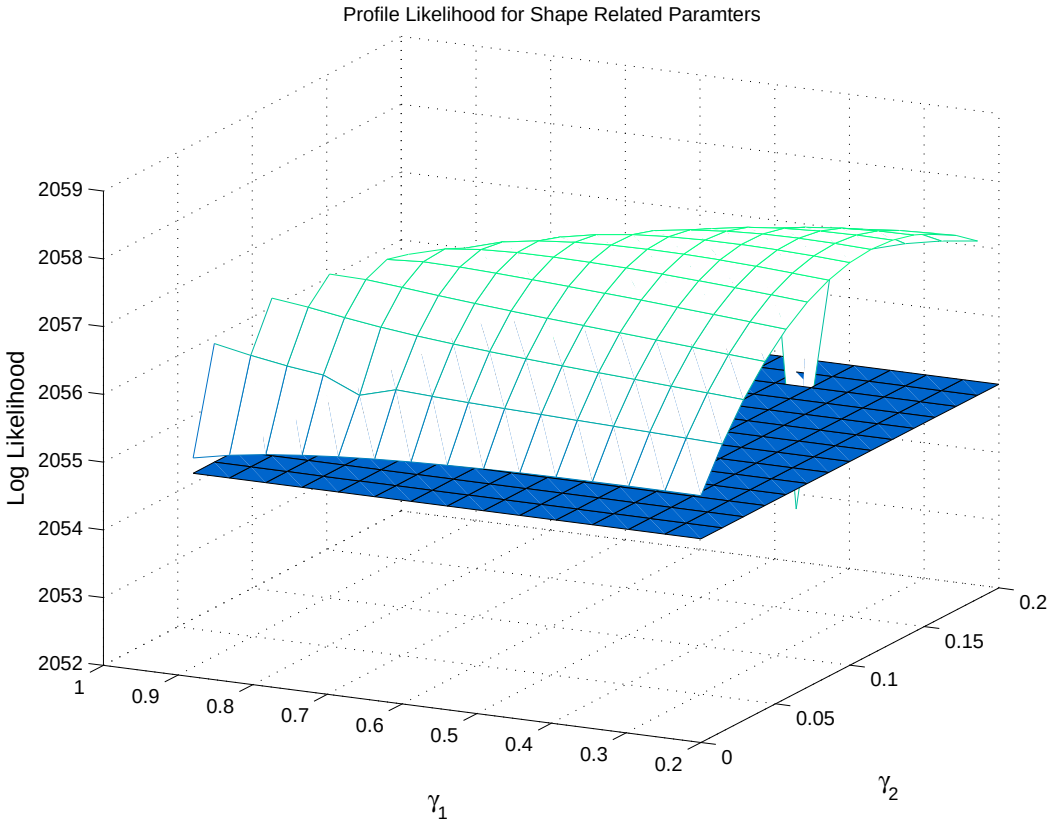

Figure 4.17: Profile likelihood for shape related parameters for data simulated from Bali's model. The curved plane is the profile likelihood surface and the dark plane is the confidence based on the 95% confidence level deviance drop.

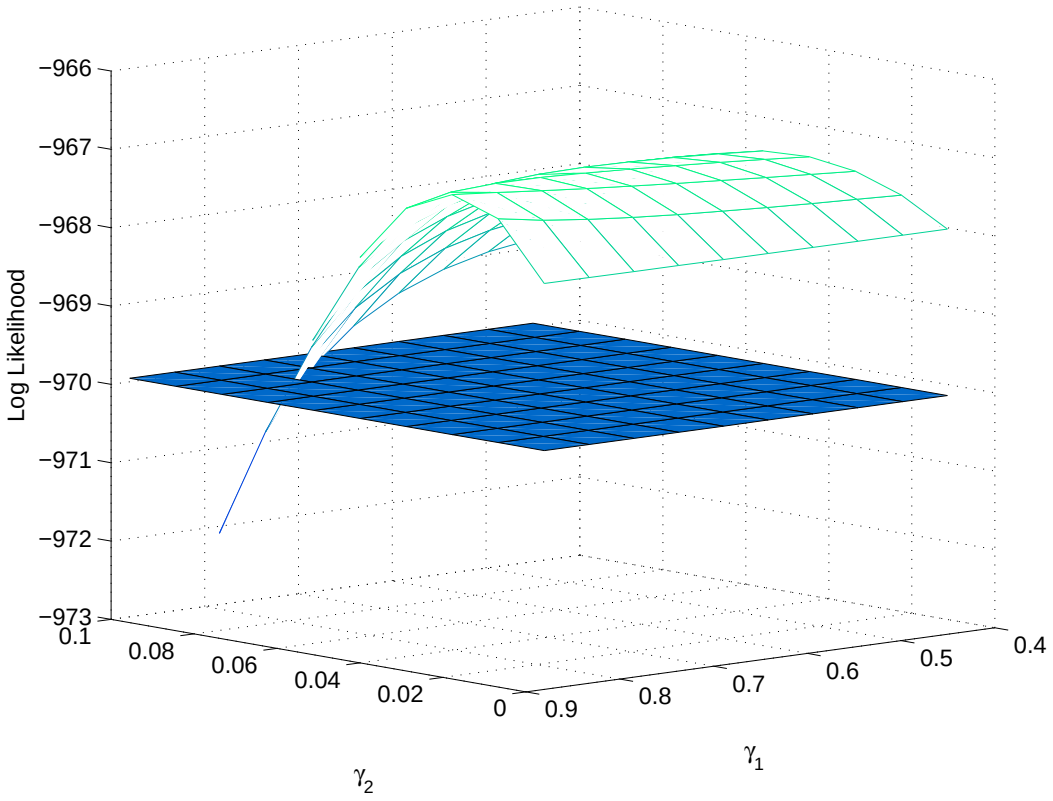

Figure 4.18: Profile likelihood for shape related parameters for data simulated from GEV-GARCH-ξ model. The curved plane is the profile likelihood surface and the dark plane is the confidence based on the 95% confidence level deviance drop.

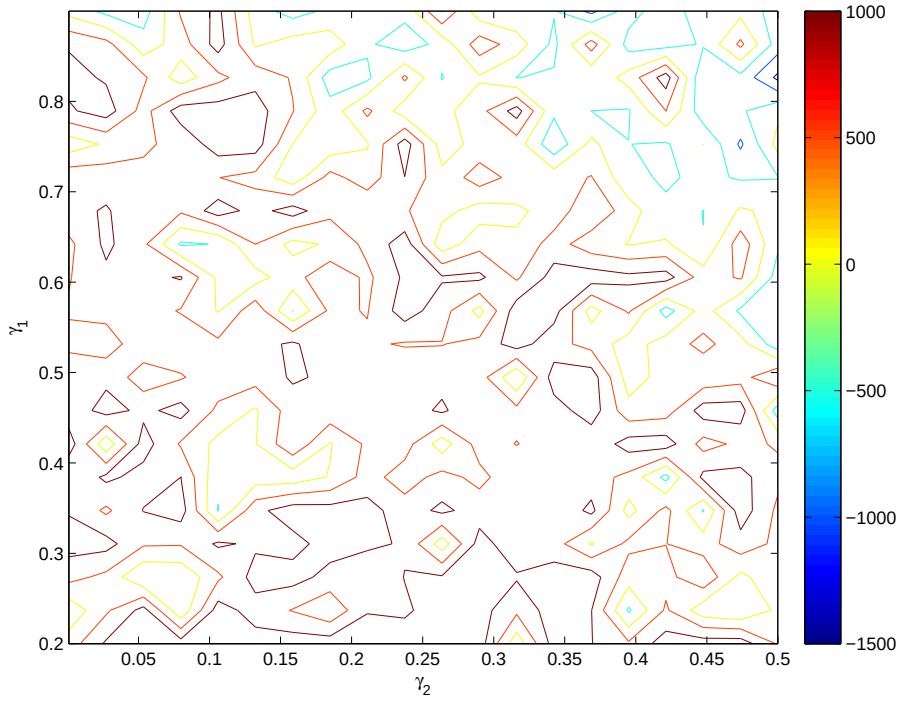

Figure 4.19: Contour plot of profile likelihood for shape related parameters of Bali and Weinbaum (2007) model for IBM stock returns.

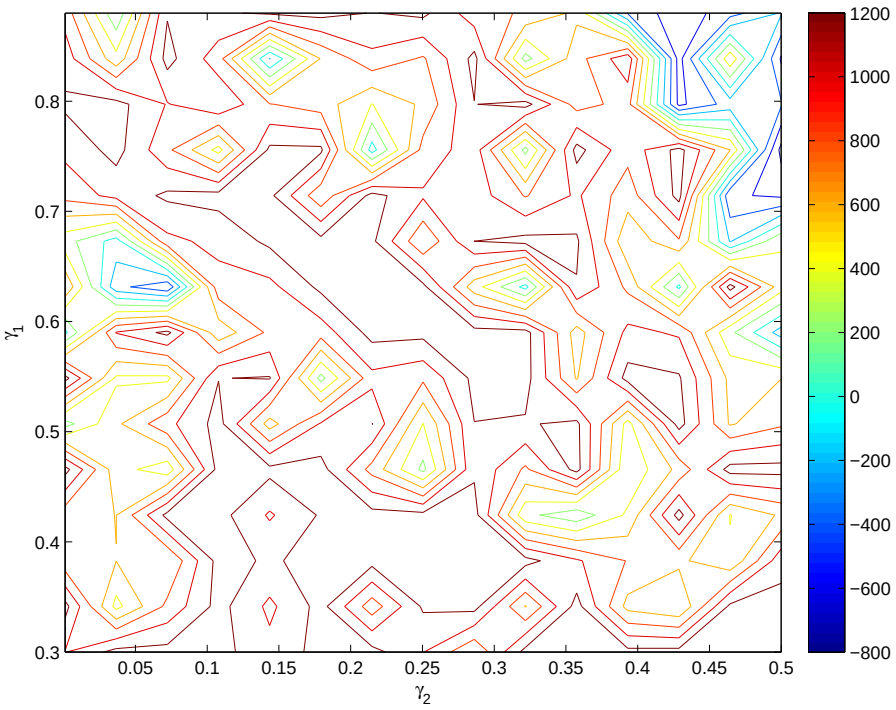

Figure 4.20: Contour plot of profile likelihood for shape related parameters of GEV-GARCH-ξ model for IBM stock returns.

Chapter 5

A Two Tail GPD Mixture Model

5.1 Introduction

5.1.1 Background

Extreme value model has been recently widely discussed within many areas as finance, economics, environment and engineering. The extreme value model describe the stochastic dynamics of a process to states with small chances of realization, and typically beyond the range of observed data. It is therefore suitable in tail related measurements. The commonly used the extreme value models include the Generalized extreme value distribution (GEV) for the block maxima or minima and the generalized pareto distribution for the excess above (below) a high (low) threshold. The typical problem related to the extreme value model is the lack of the tail information and the GPD has the advantage in the availability of the extreme observations compare to the GEV. It is therefore more preferred for the applications which is hard to collect large amount data and used more generally relatively.

The threshold selection of the GPD can be problematic in applications. First, the threshold selection has to satisfy the balance of validity of the asymptotic and variance of the estimators. The threshold must be sufficiently high to ensure the threshold excesses such that the asymptotically motivated GPD provides a reliable approximation to avoid bias. The threshold cannot be too high otherwise there is potential little sample information leading high variance on estimates. There is always trade off between these two conditions, and there is no standardized solution in the literature to find the threshold. Different threshold choice will induce a completely different tail features for some applications leading to large uncertainty due to the threshold selection. Moreover,

the classical modelling approach using the GPD treats the threshold once chosen as a fixed quantity and therefore cannot account the resultant uncertainty from the threshold choice.

The idea of threshold selection is to pick as low a threshold as possible subject to the limit model providing a reasonable approximation. Traditionally, two methods are available for this: the first method is an exploratory technique carried out prior to model estimation, e.g. using mean residual life plots, also referred to as mean excess plots in statistical literature; the second method is an assessment of the stability of parameter estimates based on the fitting of models across a range of different thresholds (Coles 2001).

Suppose a sequence of i.i.d. measurements $x_1, \ldots, x_n$ and let $x_{(1)}, \ldots, x_{(n_u)}$ represent the subset of data points that exceed a particular threshold, u, where $x_{(1)}, \ldots, x_{(n_u)}$ consist of the n_u observations that exceed u. Define threshold excesses by:

$$e_n(u) = \frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u), \quad u < \max\{x\} \tag{5.1}$$

Above a threshold u_0 at which the GPD provides a valid approximation to the excess distribution, the mean residual life plot should be approximately linear in u, see Section 2.4.2 for detailed explanation. The resulting plot, plot of the mean excess function against threshold values, is called the mean residual life plot. Confidence interval can be added to the plot based on the approximate normality of sample means.

However, the interpretation of a mean residual life plot is not always simple in practice. A change in slope of $e_n(u)$ may be often observed until a very high threshold, which however results no enough observation to make meaningful inference. Further, there is often no unique choice of u and the subjective judgement and common sense are generally involved in the decision of threshold choice. An example is shown below to explain the use of the mean residual life plot in threshold selection. Figure 5.1 shows the mean residual life plot for the Fort Collins, C.O. precipitation dataset, which can be find in *R* extReme package. The idea is to find the lowest threshold where the plot is nearly linear, taking into account the 95% confidence bounds. For the Fort Collins data, it is especially difficult to interpret, which may be because of the annual cycle (seasonality) that is being ignored here. The plot appears roughly linear from about 0.3 to 2.5 inches and the slop of the mean excess function is erratic above 2.5 inches. A range of values can be plausible for the choice of threshold.

The second method is to chose a threshold value, at which stability in the parameter estimates can be find. This method requires fitting data to the GPD distribution several

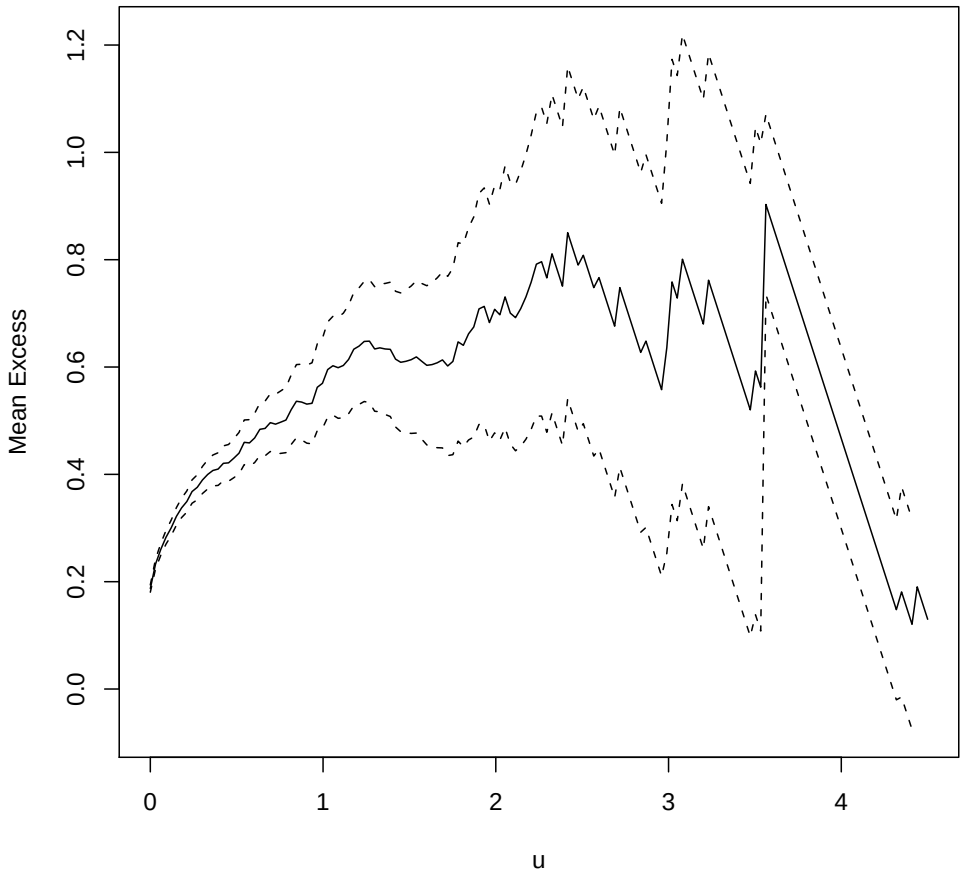

Figure 5.1: Mean residual life plot for the Fort Collins, C.O. precipitation dataset

times using different threshold. We use the Fort Collins, C.O. precipitation dataset above to illustrate the method as an example. Figure 5.2 shows the parameter estimates, $\hat{\xi}$ and $\hat{\sigma^*}$ when fitting GPD for the threshold value range $u \in (0, 1)$, which shows linearity in the residual life plot and stability of the slop of the mean excess function. Notice, the adjusted scale σ^*, $\sigma^* = \sigma_u - \xi u$, is used in the figure since that it is constant with respect to u. Again, there are several plausible values for the threshold, e.g. *0.26,0.39*, and subjective judgement is needed to pick an exact value for the threshold. Different threshold choice will result consequent later inference to be different, such as upper quantile estimation and see Section 6.4.2 for example.

Another approach is to identify the threshold choice via measuring the weights of the extreme points through robust estimations of GPD (Dupuis 1998) based on different threshold. The method fits the GPD robustly to the data using techniques based on optimal bias-robust estimates (OBRE) many times using different threshold. The robust procedure will assign weights between *0* and *1* to each data point and these weights

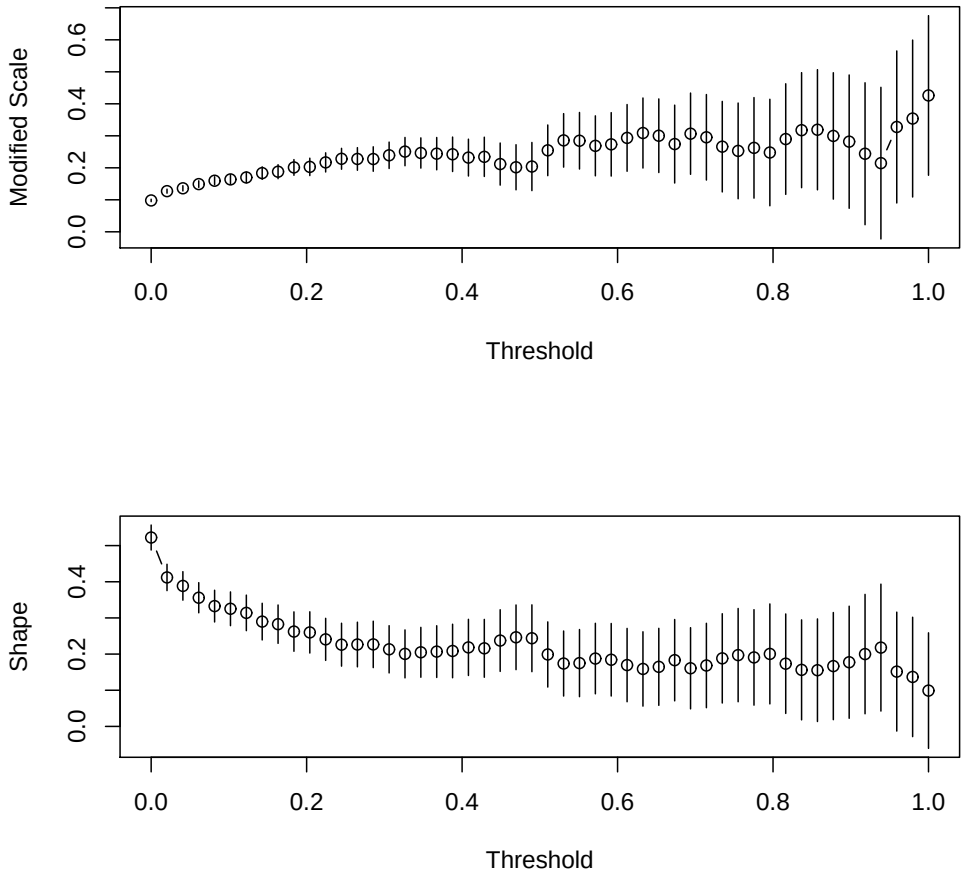

Figure 5.2: GPD fit for a range of threshold for the Fort Collins, C.O. precipitation dataset

are used to assess the validity of the GPD model for the excess over the associated threshold distribution. The OBRE will return a weight less than *1* when the norm of the appropriately standardized score function exceeds the robustness constant. Therefore, the downweights observations are not close to the working values of the parameters in the score function metric. When the observations are consistent with the current model, the weights are one and the score function is that of maximum likelihood estimates. The downweights observations are likely to be those at either extremal tails. Examining the weights of these extremes from OBRE provides the knowledge in whether our model is consistent with the extreme data and hence can be used in assess the validity of the threshold associated. The example of the OBRE for threshold choice can be found in Section 5.6. However some subjective assessment is still required to detect the potential downweights because of the lack of standard solution for it and the procedure is difficult to be applied in an automated process.

All above three approaches regard the threshold as a fixed value in latter inferences so the uncertainty associated with its selection is not accounted for and the selection can involve subjective judgements.

Another recently developed approach for threshold choice is to build a model for the entire distribution of the population whilst including the GPD above some threshold, where the threshold is typically treated as one of the parameter of the model.

Frigessi, Haug, and Rue (2002) suggest a dynamically weighted mixture model of a GPD and a light-tailed distribution for bulk of the distribution, with a smooth weight function to transition between the two distributions, to replace the usual threshold choice. Let $X_1, \ldots, X_n$ be non-negative i.i.d. random variables, the density function of the mixture distribution is given by:

$$l(x) = \frac{[1 - p(x|\theta)]\, f(x|\beta) + p(x|\theta) g(x|\xi, \sigma)}{\int_0^\infty [1 - p(x|\theta)]\, f(x|\beta) dx + \int_0^\infty p(x|\theta) g(x|\xi, \sigma) dx} \tag{5.2}$$

where, $g(x|\xi, \sigma)$ is the GPD density for the upper tail of the distribution, $f(x|\beta)$ is the density for a light-tailed distribution for bulk of the distribution, and $p(x|\theta)$ is the mixing function taking value in $(0, 1]$. In particular, they use the Weibull for the bulk of the distribution and the mixing function:

$$p(x|\theta) = \frac{1}{2} + \frac{1}{\pi} arctan\left(\frac{x - \mu}{\tau}\right), \quad \theta = (\mu, \tau), \quad \mu, \tau > 0$$

with a location μ and steepness τ^{-1}. The mixing function plays the role in which the GPD becomes predominant for the fitting of extreme observations and replace the threshold choice by the choice of mixing function. Further, non-extremes under the mixture model here contribute to inference about the tail which can result in difficulties in parameter identification.

Another similar mixture model is developed by Behrens, Lopes, and Gamerman (2004) by using a mixture of truncated Gamma for non-extremes and the GPD for one side extremes. The model also define the distributions of entire data sequence rather than exceedence over threshold and the threshold can be estimated with other model parameters. The distribution function (cumulative probability function) of their mixture model is as:

$$F(x|\eta, \xi, \sigma, u) = \begin{cases} H(x|\eta) & \text{for } x < u \\ H(u|\eta) + [1 - H(u|\eta)] G(x|\xi, \sigma, u) & \text{for } x \geq u \end{cases} \tag{5.3}$$

where, $H(x|\eta)$ denotes Gamma distribution for bulk of the distribution and $G(x|\xi, \sigma, u)$ denotes the GPD distribution for the upper tail of the distribution. As the advantage of regarding the threshold as the extra parameter of the model, the mixture GPD model is

able to account for all uncertainties of the model including that of threshold selection.

Tancredi, Anderson, and O'Hagan (2006) propose a less restrictive approach to overcome the lack of a natural model below the threshold distribution as adapting an unknown number of uniform distributions. The threshold itself is again a parameter of the model to be estimated. Let u_0 be a basic low threshold, the data (assumed to be independent and identically distributed) beyond u_0 can be modeled with the proposed mixture model with the density as:

$$f(x) = \begin{cases} (1-w)h(x|\omega^{(k)}, a^{(k)}, \alpha) & u_0 < x < \alpha \\ wg(x|\alpha, \sigma, \xi) & \alpha \leq x < \infty \end{cases} \tag{5.4}$$

where $g(x|\alpha, \sigma, \xi)$ is the GPD density with the threshold value u (unknown), and denoted by α, w can be interpreted as the probability that an observation from the distribution is greater than α conditional on being greater than the basic low threshold u_0, and $h(x|\omega^{(k)}, a^{(k)}, \alpha)$ is a piecewise constant density on the $[u_0, \alpha)$ with unknown number of k. Here,

$$h(x|\omega^{(k)}, a^{(k)}, \alpha) = \sum_{i=1}^{k} \omega_i I_{[a_i, a_{i+1})}(x)$$

where $a_1 = u_0$ and $a_{k+1} = \alpha$ and

$$\sum_{i=1}^{k} \omega_i (a_{i+1} - a_i) = 1$$

$h(x)$ can be regarded as a mixture of k uniform distributions with unknown k. The key idea of the method is the distribution below threshold is a very flexible mixture of k uniforms and the merge of the mixture of uniforms and the GPD is estimated by all the data. The model is therefore more applicable with less restriction on the distribution for the data below threshold.

These recent mixture modelling approaches mainly assume one tail of population distribution follows a GPD and a certain bulk distribution for all the data below the threshold, which assume to be bounded for the lower side of tail. These mixture models are not applicable for the applications which are with both heavily distributed tails, or applications which require reliable inferences for both tails of the population distribution simultaneously. It has been well documented that many time series in finance and economics applications having both sides heavy-tailed (not exponentially bounded) and potentially asymmetric. It is therefore natural to consider a mixture model with two sides GPD in the VaR to examine both gains and losses return extremes simultaneously. Mendes and Lopes (2004) suggest a data driven estimation process to fit a two sides

GPD mixture model by quasi-maximum likelihood on a robustly standardized data set, and the threshold will be obtained as a sub-product of the procedure. The process solves the problem in choosing threshold. However, as a result of the multistage estimation process, the uncertainty of the mixture model for the component from the threshold choice, is hard to account for. In some applications, such as Value-at-Risk (VaR) in finance, the uncertainty estimation is also important as a concern of risk control.

5.1.2 Two-tail Mixture Model

A mixture model, where both tails are GPD is developed in this chapter to capture both tails simultaneously and account for the uncertainty about the thresholds in any inferences made. More specifically, we define a mixture distribution with three components: a GPD for losses tail, a GPD for gains tail and a certain distribution for the non-extreme data between the two GPD thresholds. The distribution selected for the non-extreme data can affect estimation of the tail distribution and therefore it is necessary to choose it according to the application. The normal distribution is suggested with the interests in financial applications, due to their inherent unimodal, approximately symmetric and quadratic shape around the mode. The proposed two tail mixture GPD model can estimate both thresholds along with the other model parameter and therefore can account the all uncertainties. Clearly the extension of two GPD provides a very flexible model for capturing all forms of tail behaviour, potentially allowing for asymmetry in the distribution of two tails. The model is able to capture two sided tails distribution simultaneously, which is more applicable than a one tail mixture GPD in financial applications.

A Bayesian method of estimation is used for the mixture model as it can take the advantage of any expert prior information, which is important in tail related estimation due to the inherent sparsity of data. The reliability of the model estimation of the proposed mixture model is first tested by simulation from the proposed model. The general performance of the model is examined through fitting model to certain commonly used distributions, either symmetric or asymmetric, and evaluating the performance at estimating the high (or low) tail quantiles. Other sides analysis and comparison with other approached for extremes are also conducted including: validating the threshold estimation by the robust estimation as Dupuis (1998); the comparison in the extreme quantile estimation through different approaches; asymmetry analysis.

The chapter is organized as follows: Section 5.2.1 review the classical GPD distribution with a pre-fixed threshold. Section 5.2.2 defines the model; Section 5.3 describes the estimation method; Section 5.4 and Section 5.5 summarizes the results from the simulation studies assessing the model and estimation method performance; Section 5.6

compares the threshold estimate of the proposed GNG model with the threshold choice via the robust estimation approach of Dupuis (1998); Section 5.7 examine the asymmetry property of proposed mixture distribution, followed by summary of the chapter in Section 5.8.

5.2 Model

5.2.1 The Classical GPD

Let X be a random iid variable of GPD and represent the extremes above the threshold u, its distribution function is:

$$G(x|\xi,\sigma,u) = \begin{cases} 1-\left[1+\xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi} & \xi \neq 0 \\ 1-\exp\left(-\frac{x-u}{\sigma}\right) & \xi = 0 \end{cases} \quad x \geq u, \quad 1+\xi\left(\frac{x-u}{\sigma}\right) > 0, \sigma > 0 \tag{5.5}$$

and the density is as:

$$g_{PD}(x|\xi,\sigma,u) = \begin{cases} \frac{1}{\sigma}\left[1+\xi\left(\frac{x-u}{\sigma}\right)\right]^{-(1+1/\xi)} & \xi \neq 0, \quad \sigma > 0 \\ \frac{1}{\sigma}\exp\left(-\frac{x-u}{\sigma}\right) & \xi = 0 \end{cases} \quad x \geq u, \quad 1+\xi\frac{x-u}{\sigma} > 0, \tag{5.6}$$

where, ξ is the shape parameter, σ is the scale parameter and u is the threshold.

5.2.2 Two Tail GPD Mixture Model

A mixture model is proposed here with estimation of the thresholds along with the other model parameters. The model consider a mixture distribution of two GPD's for the upper and lower tails combined with a known distribution for the bulk of the distribution (mode). More specifically, a GPD for both tails and normal distribution for the non-extreme data between the lower and upper threshold. The proposed model has more flexible feature in extrapolation of the tails behaviors in the combinations of two tails without assuming symmetry of the two tails. In considering with the financial applications, the Normal distribution is chosen for the non-extreme data as it is more appropriate for the financial application to their inherent unimodal, approximately symmetric and quadratic shape around the mode. The following paper will use GNG (GPD-Normal-GPD) to denote the proposed mixture distribution. Let X be a independent and identically distributed random variables from a GNG distribution. The

distribution function of mixture model , $P(X < x) = F(x)$ is as:

$$\begin{aligned} F(x|m,s,\xi_r,\sigma_r,u_r,\xi_l,\sigma_l,u_l) = &\{\Phi(u_l|m,s)[1-G(-x|\xi_l,\sigma_l,-u_l)]\}\, I_{(-\infty,u_l]}(x) \\ &+ \Phi(x|m,s) I_{(u_l,u_r)}(x) \\ &+ \{\Phi(u_r|m,s) + [1-\Phi(u_r|m,s)]G(x|\xi_r,\sigma_r,u_r)\}\, I_{[u_r,\infty)} \end{aligned} \tag{5.7}$$

where, $\Phi(x)$ is the normal cumulative distribution function with mean m and variance s^2 given by:

$$\Phi(x|m,s) = \frac{1}{s\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{-\frac{(t-m)^2}{2s^2}\right\} dt, x \in R \tag{5.8}$$

and $G(x|\xi,\sigma,u)$ is the distribution function of GPD defined as Equation 5.5. The subscript "$_l$" denote the left tail and "$_r$" denote the right tail of the distribution. For example, u_l is the left(low) threshold for the mixture distribution with corresponding GPD parameters σ_l and ξ_l. $I(x)$ is the indicator function reflecting the subset of range support of x. Theorem 5.2.1 shows that Equation 5.5 is a valid probability function. Therefore, the parameter vector of the model is $\theta = (m, s, u_r, \xi_r, \sigma_r, u_l, \xi_l, \sigma_l)$. The likelihood of GNG is given by:

$$\begin{aligned} L(\theta;x) = &\prod_{(-\infty,u_l]} \Phi(u_l|m,s) \left\{\frac{1}{\sigma_l}\left[1+\xi_l\left(\frac{u_l-x_i}{\sigma_l}\right)\right]_+^{-\frac{1+\xi_l}{\xi_l}}\right\} \prod_{(u_l,u_r)} \phi(x_i|m,s) \\ &\prod_{[u_r,\infty)} [1-\Phi(u_r|m,s)] \left\{\frac{1}{\sigma_r}\left[1+\xi_r\left(\frac{x_i-u_r}{\sigma_r}\right)\right]_+^{-\frac{1+\xi_r}{\xi_r}}\right\} \qquad \xi \neq 0 \end{aligned} \tag{5.9}$$

for non-zero shape, and when we replace the density of GPD as the case in Equation 5.6 when $\xi = 0$, the likelihood for a zero shape then can be defined as:

$$\begin{aligned} L(\theta;x) = &\prod_{(-\infty,u_l]} \Phi(u_l|m,s) \frac{1}{\sigma_l} \exp\left(-\frac{u_l-x_i}{\sigma_l}\right) \prod_{(u_l,u_r)} \phi(x_i|m,s) \\ &\prod_{[u_r,\infty)} [1-\Phi(u_r|m,s)] \frac{1}{\sigma_r} \exp\left(-\frac{x_i-u_r}{\sigma_r}\right) \qquad \xi = 0 \end{aligned} \tag{5.10}$$

Theorem 5.2.1. A valid probability function of the two tail GPD mixture model

The specified distribution function (as Equation 5.7) of two tail mixture model is a valid distribution function.

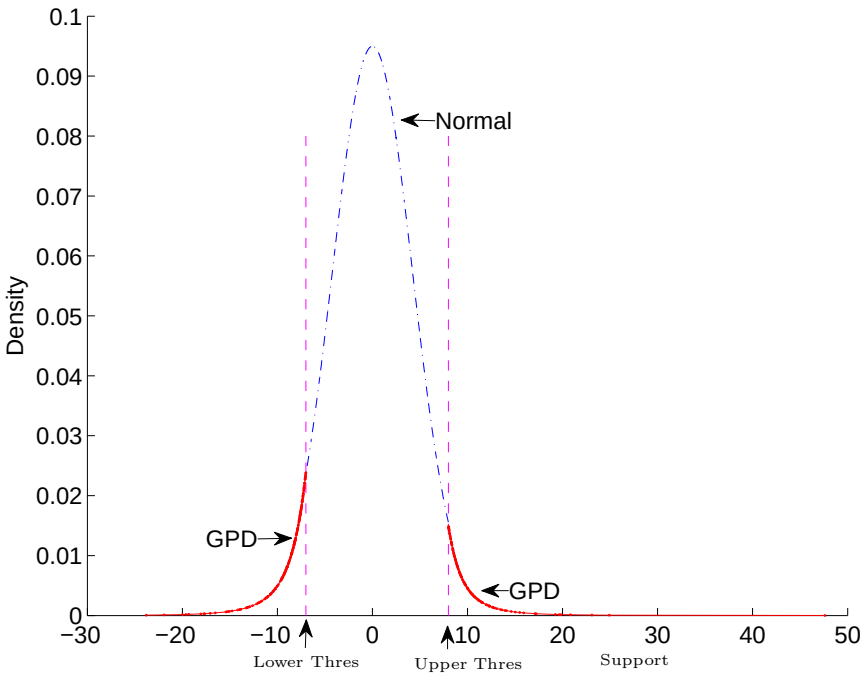

Figure 5.3: Example of the two tailed GNG (GPD-Normal-GPD) model with smooth density. The two vertical dash line represent the threshold cut off points for the two GPD distributions.

In the thesis, the distribution function of the mixture model is given by:

$$\begin{aligned} F(x|m,s,\xi_r,\sigma_r,u_r,\xi_l,\sigma_l,u_l) = &\{\Phi(u_l|m,s)[1-G(-x|\xi_l,\sigma_l,-u_l)]\}\, I_{(-\infty,u_l]}(x) \\ &+ \Phi(x|m,s) I_{(u_l,u_r)}(x) \\ &+ \{\Phi(u_r|m,s) + [1-\Phi(u_r|m,s)]G(x|\xi_r,\sigma_r,u_r)\}\, I_{[u_r,\infty)} \end{aligned}$$

where, $\Phi(x)$ is the normal cumulative distribution function with mean m and variance s^2 and $G(x|\xi,\sigma,u)$ is the distribution function of GPD. The subscript "$_l$" denote the left tail and "$_r$" denote the right tail of the distribution. For example, u_l is the left(low) threshold for the mixture distribution with corresponding GPD parameters σ_l and ξ_l. $I(x)$ is the indicator function reflecting the subset of range support of x.

Proof. To be a valid cumulative distribution function, $F(x)$ must satisfy the following conditions:

$$\lim_{x\to-\infty} F(x) = 0$$

$$\lim_{x\to\infty} F(x) = 1$$

and, $F(x)$ a non-decreasing function for $x \in (-\infty,\infty)$.

Let $S = x \in (-\infty, \infty)$ be the support of X and it can be divided into three components as $S_1 = \{x \in (-\infty, u_l]\}$, $S_2 = \{x \in (u_l, u_r)\}$ and $S_3 = \{x \in [u_r, \infty)\}$. Since they are mutually exclusive, $P(S) = P(S_1) + P(S_2) + P(S_3) = 1$.
Let the distribution function of mixture model be .
The bulk distribution can be any known distribution function. In the mixture model above, we choose normal distribution denoted by $\Phi(x)$ and the support is S_2.

A valid CDF has upper limit as $\lim_{x\to\infty} \Phi(x) = 1$.
So, the sum of the three components F(x) is as

$$\begin{aligned}\lim_{x\to\infty} F(x) &= P(X \le u_l) + P(u_l < X < u_r) + P(X \ge u_r) \\ &= \Phi(u_l) + [\Phi(u_r) - \Phi(u_l)] + [1 - \Phi(u_r)] \\ &= 1\end{aligned}$$

Use $G(x|u)$ denote the conditional distribution function of GPD for a given threshold value u and $x \ge u$.

- For $x \in S_1$ - lower tail:

  $$P(X \le u_l) = \Phi(u_l)$$

  and

  $$P(x|x \le u_l) = 1 - P(-x| - x \ge -u_l) = 1 - G(-x| - u_l)$$

  Therefore

  $$P(x \cap x \le u_l) = F(x)_{x\in(-\infty,u_l]} = [1 - G(-x| - u_l)]\Phi(u_l)$$

  As an valid distribution function of GPD,

  $$\lim_{-x\to\infty} G(-x| - u_l) = 1$$

  Therefore

  $$\begin{aligned}\lim_{x\to-\infty} F(x) &= [1 - \lim_{-x\to\infty} G(-x| - u_l)]\Phi(u_l) \\ &= (1 - 1)\Phi(u_l) \\ &= 0\end{aligned}$$

  Also, $G(-x| - u_l)$ decrease as x increase. Therefore $F(x)$, for $x \in S_1$ is a non-decreasing function.

- For bulk distribution

$$\begin{aligned} F(x)_{x\in(u_l,u_r)} &= P(X \le u_l) + P(u_l < X < x) \\ &= \Phi(u_l) + [\Phi(x) - \Phi(u_l)] \\ &= \Phi(x) \end{aligned}$$

  Since $\Phi(x)$ is a non-decreasing function, $F(x)$, for $x \in S_2$, is a non-decreasing function.

- Similarly, for upper tail:

$$P(X \ge u_r) = 1 - \Phi(u_r)$$

  and

$$P(x|x \ge u_r) = P(x|x \ge u_r) = G(x|u_r)$$

  Therefore

$$P(x \cap x \ge u_r) = G(x|u_r)[1 - \Phi(u_r)]$$

$$\begin{aligned} F(x)_{x\in[u_r,\infty)} &= P(X < u_r) + P(x \cap x \ge u_r) \\ &= \Phi(u_r) + G(x|u_r)[1 - \Phi(u_r)] \end{aligned}$$

  Since $G(x|u_r)$ is a non-decreasing function, $F(x)$, for $x \in S_3$, is a non-decreasing function.

We use indicator function $I_{(-\infty,u_l]}(x)$ to denote lower tail, $I_{(u_l,u_r)}(x)$ to denote bulk and $I_{[u_r,\infty)}(x)$ to denote upper tail.
So, for the support $S = \{x \in (-\infty, \infty)\}$, the distribution function $F(x)$ is as

$$\begin{aligned} F(x|m, s, \xi_r, \sigma_r, u_r, \xi_l, \sigma_l, u_l) = &\{\Phi(u_l|m, s)[1 - G(-x|\xi_l, \sigma_l, -u_l)]\} I_{(-\infty,u_l]}(x) \\ &+ \Phi(x|m, s) I_{(u_l,u_r)}(x) \\ &+ \{\Phi(u_r|m, s) + [1 - \Phi(u_r|m, s)]G(x|\xi_r, \sigma_r, u_r)\} I_{[u_r,\infty)} \end{aligned}$$

That all three components of $F(x)$ are non-decreasing function leads $F(x)$ a non-decreasing function, for $x \in (-\infty, \infty)$.

To summarize above, we have shown the three conditions for being a valid CDF are satisfied, namely:

$$\lim_{x\to-\infty} F(x) = 0$$

$$\lim_{x\to\infty} F(x) = 1$$

and, $F(x)$ a non-decreasing function for $x \in (-\infty, \infty)$. □

Figure 5.3 and Figure 5.4 gives examples of GNG distributions. The two vertical dash lines represent the thresholds for the two tail GPD distributions. The density plot is smooth at the value two distribution meet together for example by Figure 5.3. But it is not always the case and the density of the mixture distribution may have a discontinuity at the cut off points which indicates the density jump of Normal and GPD distributions as example showed by Figure 5.4. However, frequently in applications the density is close to continuous at the thresholds and any lack of continuity is typically of no concern if interests are only the extremes.

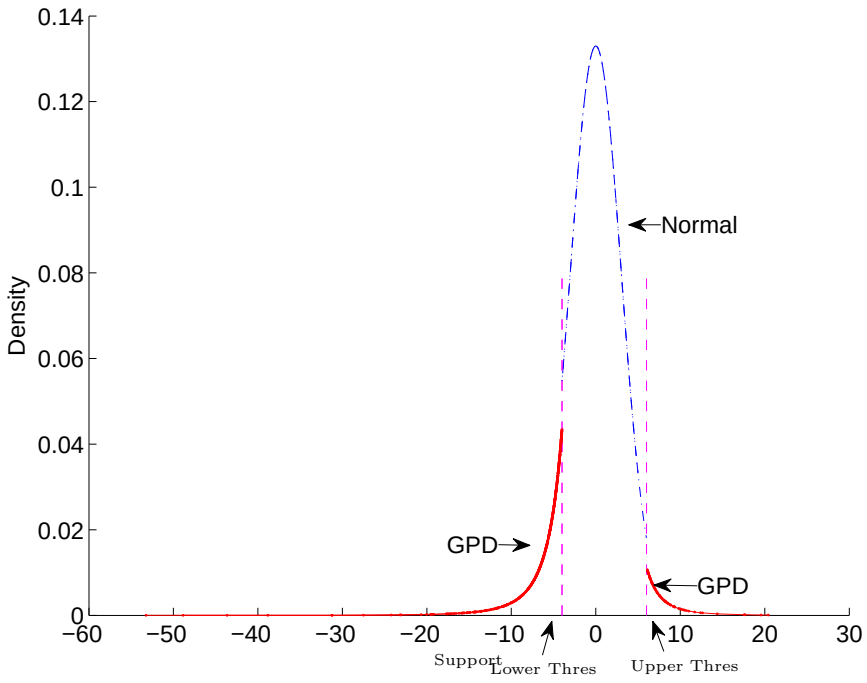

Figure 5.4: Example of the two tailed GNG (GPD-Normal-GPD) model without smooth density. The two vertical dash line represent the threshold cut off points for the two GPD distributions.

Although the extra parameters of the proposed mixture model increase the complexity for statistical inference the model has numerous advantages:

1. The model can avoid the relatively subjective threshold selection by explicitly define the thresholds as an extra parameters, which can be estimated using the standard inference techniques;

2. The mixture model can capture the uncertainty of estimating the threshold directly in estimation of the model parameters, which is more complex for the traditional fixed threshold approach;

3. The model is able to extrapolate two side tail distributions simultaneously, which may be more applicable than a one side tail model for financial applications;

4. The proposed mixture model has the flexibility in dealing with variety of the distribution with or without the symmetry feature by allowing both tails to follow GPD distribution.

5.3 Bayesian Inference for The GNG Model

The typical problem related to the estimation of extremal type models is the lack of the extremal observations. The inherent sparsity of extremes promotes interest in being able use information from all sources for inference purpose, e.g. expert knowledge. The Bayesian inference is therefore used in estimating the parameter of the mixture model in order to combine the prior information along with the sample data. Markov chain Monte Carlo (MCMC) has been used to obtain posterior distributions (Green 1995; Gamerman 1997; Robert 1999). The details of the estimation method is described in the following section.

5.3.1 Prior Distributions

The parameter vector $\theta = (m, s, u_r, \xi_r, \sigma_r, u_l, \xi_l, \sigma_l)$ can be decomposed into three components $\theta_1 = (m, s)$,$\theta_3 = (\xi_r, \sigma_r, \xi_l, \sigma_l)$, and $\theta_2 = (u_r, u_l)$,which are associated with the normal, GPD parameters, and the thresholds. We consider Bayesian estimation assuming that we do not have appreciable prior information and this means the prior distribution of the parameters will be relative diffuse. However, Bayesian estimation can always take the advantages when the expert provision available with a more informative prior in applications.

Prior for the GPD parameters

Following Coles and Tawn (1996) the GPD priors are specified on the quantile differences since that the expert prior beliefs on extremal behaviors could be hard to express directly in terms of GPD parameters adequately, and which is a fairly standard approach in the extremes in the literature. The benefit of eliciting prior using the quantile differences is that they can also account for the negative dependence between the shape ξ and scale σ parameters. The prior distribution of the difference in the quantiles is specified in terms of Gamma distribution.

By inverse the distribution function, the quantile (conditional on $x > u$) of $1-p$ (or return level with return period $1/p$, q_p) for the GPD upper tail can be expressed as Equation 2.6. The quantile differences can be calculated as $d_{q_i} = q_{p_i} - q_{p_{i-1}}, i = 1, 2$ with $q_{p_1} < q_{p_2}$, and q_{p_0} is the physical lower end point of the underlying variable for upper tail. The quantile difference for lower tail can be calculated as $d_{q_i} = q_{p_{i-1}} - q_{p_i}, i = 1, 2$ with $q_{p_1} > q_{p_2}$, and q_{p_0} is the physical upper end point of the underlying variable for lower tail.

Assuming the quantile differences a Gamma distribution $d_{q_i} \sim Ga(a_i, b_i)$, $p_1 = 0.1, p_2 = 0.01$ and $q_{p_0} = 0$ for the standardized data, the prior for upper tail is defined by:

$$\pi(\xi, \sigma) \propto Jf(q_{p_1}, q_{p_2})$$

$$f(q_{p_1}, q_{p_2}) \propto q_{p_1}^{a_1-1} e^{-b_1 q_{p_1}} (q_{p_2} - q_{p_1})^{a_2-1} e^{-b_2(q_{p_2} - q_{p_1})}$$

where J is the Jacobian transformation term and f is the joint distribution of the quantile differences. Therefore, the prior for the upper tail GPD parameters ξ and σ is:

$$\begin{aligned}\pi(\sigma, \xi) \propto & \exp\left\{-b_1\left[u + \frac{\sigma}{\xi}(p_1^{-\xi} - 1)\right]\right\}\left[u + \frac{\sigma}{\xi}(p_1^{-\xi} - 1)\right]^{a_1-1} \\ & \times \exp\left\{-b_2\left[\frac{\sigma}{\xi}(p_2^{-\xi} - p_1^{-\xi})\right]\right\}\left[\frac{\sigma}{\xi}(p_2^{-\xi} - p_1^{-\xi})\right]^{a_2-1} \\ & \times \left|\frac{\sigma}{\xi^2}\left[(p_1 p_2)^{-\xi}(\log p_1 - \log p_2) + p_2^{-\xi}\log p_2 - p_1^{-\xi}\log p_1\right]\right|\end{aligned} \tag{5.11}$$

The prior for lower tail GPD parameters is defined in the same way as the upper tail of the GPD.

Prior for the thresholds

A truncated normal distribution with hyper-parameter $(m_{u_r}, s_{u_r}, l_{u_r})$ is used as the prior distribution for threshold of upper tail, which are truncated at the minimum and maximum of the sample data respectively, due to Behrens, Lopes, and Gamerman (2004). The density of it is as:

$$\pi(u_r | m_{u_r}, s_{u_r}, l_{u_r}) = \frac{1}{\sqrt{2\pi s_{u_r}^2}} \frac{\exp\left[\frac{-1}{2}(\frac{u_r - m_{u_r}}{s_{u_r}})^2\right]}{\Phi\left[\pm(\frac{l_{u_r} - m_{u_r}}{s_{u_r}})\right]} \quad u_r \in (l_{u_r}, \infty) \tag{5.12}$$

$$\pi(u_r | m_{u_r}, s_{u_r}, l_{u_r}) \propto \exp\left[\frac{-1}{2}(\frac{u_r - m_{u_r}}{s_{u_r}})^2\right] \tag{5.13}$$

here, m_{u_r} is the location, s_{u_r} is the standard deviation and l_{u_r} is the upper limit for the upper threshold. The prior for the lower threshold is defined in the same way as:

$$\pi(u_l|m_{u_l}, s_{u_l}, l_{u_l}) \propto \exp\left[\frac{-1}{2}(\frac{u_l - m_{u_l}}{s_{u_l}})^2\right] \quad u_l \in (-\infty, l_{u_l}) \tag{5.14}$$

here, m_{u_l} is the location, s_{u_l} is the standard deviation and l_{u_l} is the lower limit for the lower threshold.

Prior for Normal Bulk Parameters

We choose a normal prior $\phi(m_m, m_s)$ for the normal location m and a gamma prior $g(\alpha, \beta)$ for the normal scale parameter s, assuming m and s are independent. These priors densities are defined as:

$$\pi(m|m_m, s_m) \propto \exp\left[-\frac{(m - m_m)^2}{2s_m^2}\right] \tag{5.15}$$

$$\pi(s|\alpha, \beta) \propto s^{\alpha-1}e^{\frac{-s}{\beta}} \tag{5.16}$$

5.3.2 Posterior Distribution

Given the prior distribution $\pi(\theta)$ and the likelihood $L(x|\theta)$, the posterior is $p(\theta|x) \propto \pi(\theta)l(x|\theta)$. The log posterior of the model is therefore defined as:

$$
\begin{aligned}
\log p(\theta|x) =& K + \sum_{i=1}^{n} I(u_l < x_i < u_r) \left[-\log s - \frac{1}{2}(\frac{x_i - m}{s})^2 \right] + \sum_{i=1}^{n} I(x_i \geq u_r) \log\left[1 - \Phi(u_r|m,s)\right] \\
&+ \sum_{i=1}^{n} I(x_i \geq u_r) \left\{ -\log \sigma_r - \frac{1+\xi_r}{\xi_r} \log\left[1 + \frac{\xi_r(x_i - u_r)}{\sigma_r}\right] \right\} \\
&+ \sum_{i=1}^{n} I(x_i \leq (u_l) \log\left[\Phi(u_l|m,s)\right] \\
&+ \sum_{i=1}^{n} I(x_i \leq u_l) \left\{ -\log \sigma_l - \frac{1+\xi_l}{\xi_l} \log\left[1 + \frac{\xi_l(u_l - x_i)}{\sigma_l}\right] \right\} - \left[\frac{1}{2s_m^2}(m - m_m)^2 \right] \\
&+ (\alpha - 1)\log(s) - \frac{s}{\beta} - \frac{1}{2}(\frac{u_r - m_{u_r}}{s_{u_r}})^2 - \frac{1}{2}(\frac{u_l - m_{u_l}}{s_{u_l}})^2 \\
&- b_{1_r} \left[u_r + \frac{\sigma_r}{\xi_r}(p_{1_r}^{-\xi_r} - 1) \right] + (a_{1_r} - 1)\log\left[u_r + \frac{\sigma_r}{\xi_r}(p_{1_r}^{-\xi_r} - 1) \right] \\
&- b_{2r} \left[\frac{\sigma_r}{\xi_r}(p_{2_r}^{-\xi_r} - p_{1_r}^{-\xi_r}) \right] + (a_{2_r} - 1)\log\left[\frac{\sigma_r}{\xi_r}(p_{2_r}^{-\xi_r} - p_{1_r}^{-\xi_r}) \right] \\
&+ \log\left| \frac{\sigma_r}{\xi_r^2} \left[(p_{1_r}p_{2_r})^{-\xi_r}(\log p_{1_r} - \log p_{2_r}) + p_{2_r}^{-\xi_r}\log p_{2_r} - p_{1_r}^{-\xi_r}\log p_{1_r} \right] \right| \\
&- b_{1_l} \left[u_l + \frac{\sigma_l}{\xi_l}(p_{1_l}^{-\xi_l} - 1) \right] + (a_{1_l} - 1)\log\left[u_l + \frac{\sigma_l}{\xi_l}(p_{1_l}^{-\xi_l} - 1) \right] \\
&- b_{2_l} \left[\frac{\sigma_l}{\xi_l}(p_{2_l}^{-\xi_l} - p_{1_l}^{-\xi_l}) \right] + (a_{2_l} - 1)\log\left[\frac{\sigma_l}{\xi_l}(p_{2_l}^{-\xi_l} - p_{1_l}^{-\xi_l}) \right] \\
&+ \log\left| \frac{\sigma_l}{\xi_l^2} \left[(p_{1_l}p_{2_l})^{-\xi_l}(\log p_{1_l} - \log p_{2_l}) + p_{2_l}^{-\xi_l}\log p_{2_l} - p_{1_l}^{-\xi_l}\log p_{1_l} \right] \right| \\
&\text{when} \xi \neq 0
\end{aligned}
\tag{5.17}
$$

Here, the K is the normalizing constant and the posterior is only for mixture distribution when $\xi \neq 0$. In the case where $\xi = 0$, the posterior can be obtained by replacing above function with the likelihood and prior of zero shape for either one tail or both tails.

5.3.3 MCMC Procedure

Posterior samples are obtained through Markov chain Monte Carlo methods (MCMC). Posteriors sampling is through Metropolis-Hastings (M-H) algorithm as its advantage of being free of functional form when the posterior distribution function is not a proper probability function. A random walk M-H algorithm is used for the method of sampling. The possibility of $\xi = 0$ has been included into the algorithm. The M-H algorithm

consists two stages: First, a draw from the proposal density is obtained; Second, the draw is either retained or rejected. The M-H algorithm is very similar as Section 4.4.3 and will not be detailed here for brevity.

The convergence of the iterative simulation of posterior is assessed by monitoring multiple simulated sequence with dispersed starting values as suggested by Gelman, Carlin, Stern, and Rubin (2004). We can estimate the marginal posterior variance by a weighted average of the between (B) and within (W) sequence variances for each of parameter estimator as $\hat{var}(\theta_i|x) = \frac{n-1}{n}W_{var} + \frac{1}{n}B_{var}$, here n indicates the length of simulation. This is commonly used to the classical variance estimate with cluster sampling. For a finite n, the within variance will under-estimate $var(\theta_i|x)$ because the individual chain has not reach the entire target distribution. The expectation of W approaches $var(\theta_i)$ as the limit $n \to \infty$. We assess the convergence by monitoring that the scale of the current posterior distribution for θ_i might be reduced if the simulation continue in the limit $n \to \infty$. The potential scale reduction is defined as the ratio of the marginal posterior variance and within variance as $\hat{R} = \sqrt{\hat{var}(\theta_i|x)/W}$, which should decline to 1 as the chain length goes to infinity.

Only the converged chain is regarded as an approximation sample from posterior distributions. The second half of the chain is used as the posterior distribution and the estimated parameter is calculated as the mean of highest posterior density interval (HPD) of the posteriors. In the research, MCMC updated component by component of the parameter vector at iteration t in terms of the importance order to the distribution, as ξ, σ, u for both tail and then normal related m, s. The full details of algorithm is included in Appendix B.

5.4 Simulation Study I - Simulation from the GNG Model

Various simulation studies were undertaken to assess the performance of the GNG model and estimation method for various applications. The first simulation study was designed to assess the performance of the Bayesian inference approach, via an application to data simulated directly from the model with known parameter values. We first illustrate the posterior properties of the model through one simulation in Section 5.4.1 and then address the accuracy of the estimation through multiple simulations in Section 5.4.2.

5.4.1 Simulated Sample

A sample is simulated from the GNG model using the parameter value $\theta = (\mu_m = 0, s_m = 4.2, u_r = 6, \xi_r = 0.3, \sigma_r = 2.2, u_l = -5, \xi_l = 0.2, \sigma_l = 2.5)$ with a sample size 3000, with approximately 335 observations in the lower tail below u_l and 230 in the upper tail above u_r. The distribution of this simulated dataset from the GNG model is presented as Figure 5.5 left. The parameters are chosen to give two reasonably heavy tails, and a near continuous density function at the thresholds. Figure 5.5 shows the posterior predictive density and corresponding cumulative distribution function (CDF).

Figure 5.5.1 *Fitted Density*

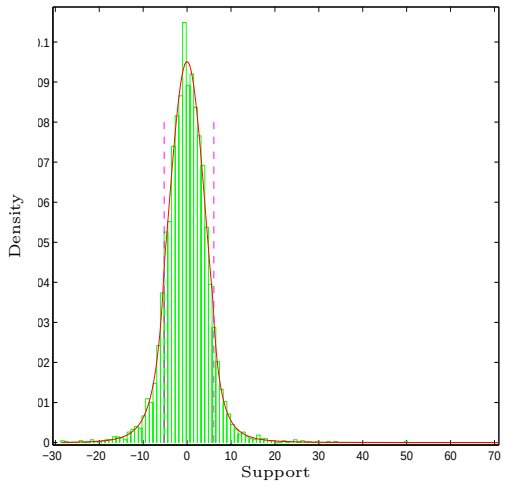

Figure 5.5.2 *CDF and Tail Return Levels*

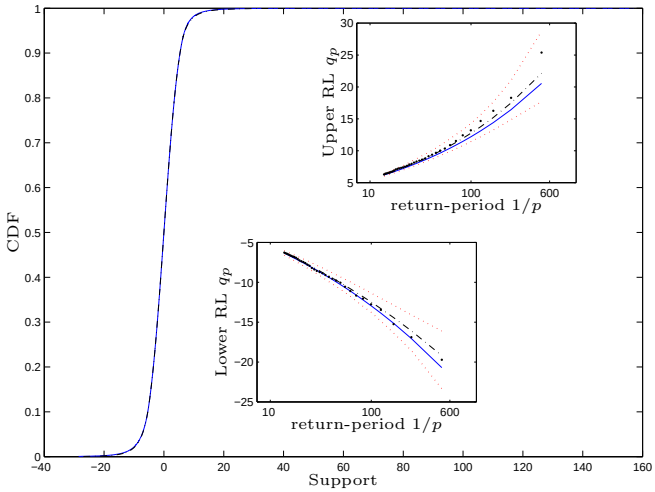

Figure 5.5: Example dataset from the GNG model with parameter set $\theta = (\mu_m = 0, s_m = 4.2, u_r = 6, \xi_r = 0.3, \sigma_r = 2.2, u_l = -5, \xi_l = 0.2, \sigma_l = 2.5)$. (1) is the density with fitted model and (2) gives the fitted CDF with excerpts showing the fit in the tails in more detail. In (1) the estimated thresholds are shown by vertical dashed lines and the posterior predictive density estimate is shown by the solid line. The true CDF and return level are denoted by solid line, sample values are presented by dots, and posterior predictive estimates are shown by dashed lines in (2).

Table 5.1 reports the parameter estimation results for the simulation along with the credible interval of posteriors. It is clear from Figure 5.5 and Table 5.1 that the Bayesian inferences are reliable, with the fitted model providing a very good fit. Notice that point estimates of the parameters are close to the true values which are well within the 95% credible intervals. The return level plot shows the estimated value are very close to true return level for both tails and the discrepancy is due to the sample variation. Particularly, the return level plot indicates a precise extreme quantile estimate even for the quantiles out of the scope of the observations. The credible interval of return level contain all sample points and true values.

Figure 5.6 shows the posterior distribution for all the parameter estimates. Generally

Table 5.1: Results from Bayesian inference for single simulated dataset with sample size n=3000 from from the GNG model. The true parameter values (True) and estimated parameters (Estimated) using the mean of the MCMC samples within the 95% highest posterior credible interval are shown.

Parameters	m	s	u_r	ξ_r	σ_r	u_l	ξ_l	σ_l
True	0	4.2	6	0.3	2.2	-5	0.2	2.5
Estimated	-0.0011	4.1960	6.0986	0.3015	2.4318	-5.2267	0.1532	2.5251
Lower CI	-0.1179	4.1022	5.6346	0.2038	2.0456	-5.7949	0.0583	2.1555
Upper CI	0.1108	4.2827	6.6472	0.4055	2.8273	-4.6149	0.2572	2.9343

the thresholds are the most difficult parameter to estimate since it can take a wide range of values and still provide an adequate fit. The posterior distribution of thresholds can be spread and possibly multimodal. The normal-related parameters have a relatively smaller variance compare to the GPD related parameters, since much more information contained in the data about the bulk distribution compare to the tails. Less variation of posterior can be expected with larger sample size and more informative prior information when appreciable export information is available.

Figure 5.7 shows the estimated relationship of the GPD parameters for both tails as contours from their joint posterior distribution. As expected, the ξ and σ are negatively correlated, and the shape parameter also appears independent of the threshold which is as expected. The scale parameters are also linearly related to the threshold as we would expect. The shape parameter posterior density is positively skewed and the scale parameter is slightly positively skewed as we would expect. The threshold posterior densities also appears to be approximately normally distributed.

Figure 5.8 displays the relationship of GPD paired parameters again after adjust the scale posteriors by $\sigma^* = \sigma - \xi(u - \mu)$. Where σ^* is adjusted sigma and μ is the GEV location and is regarded as *0* for this figure. The adjusted scale is no longer correlated to the threshold value as expected. And the negative correlation between the shape and scale is more stronger than before.

Posterior predictive checks are important diagnostics for MCMC methods to assess their performance. The basic idea is to compare a specified test quantity and an appropriate predictive distribution from posterior replications. A large discrepancy between them would indicate that the model is not a good fit to the data. The obvious quantities for diagnostics are the quantiles of the simulated sample. The posterior predictive quantile distributions are shown in Figure 5.9, along with the true quantiles and direct sample estimates of the quantiles. The sample and true quantiles are well within the 95% credible intervals and are located near the mode of the posterior predictive quantile distributions. You will notice that as the tail quantiles get more extreme the posterior predictive distri-

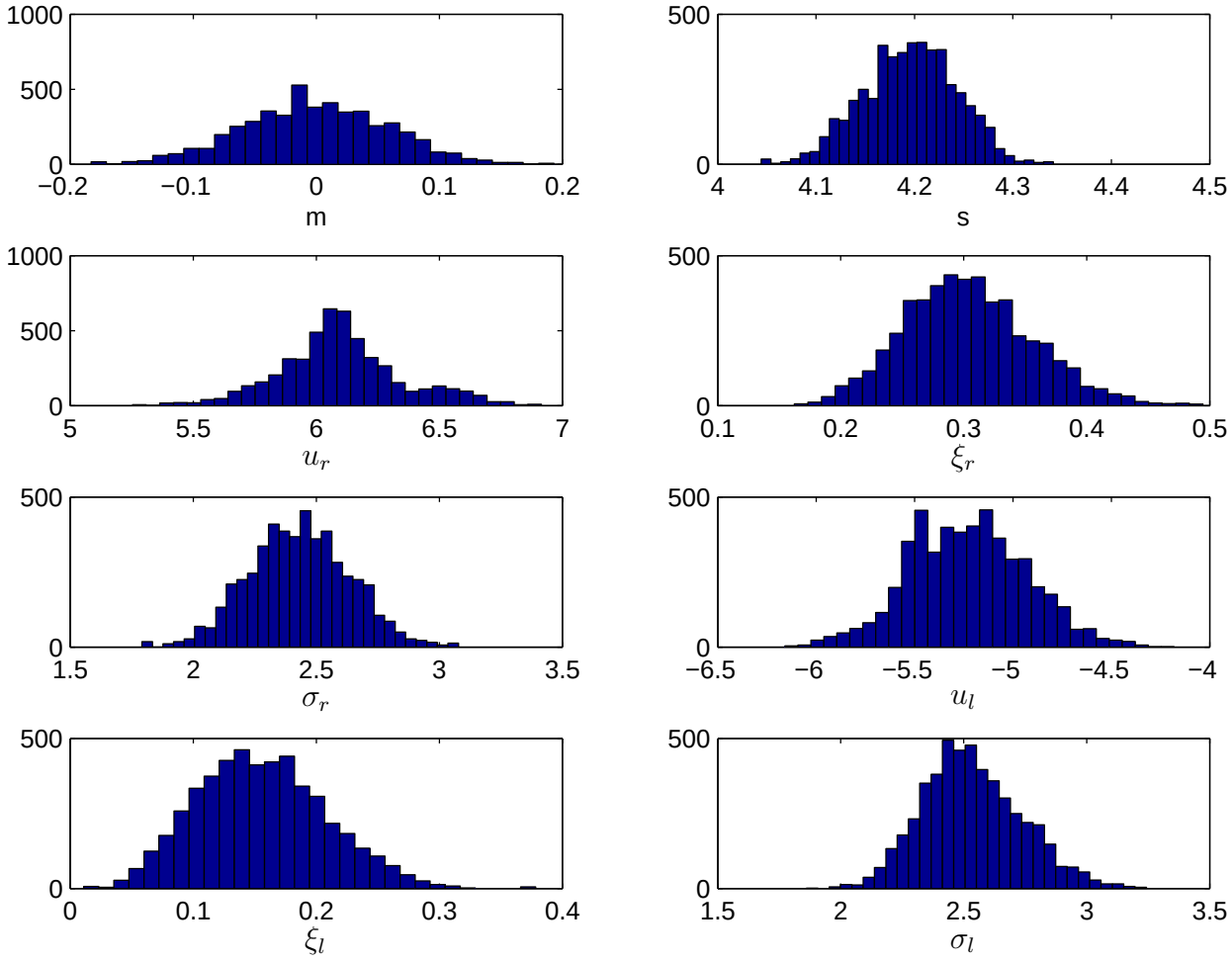

Figure 5.6: Marginal posterior distributions. ξ, σ, u are the shape, scale and threshold of GPD. The upper tail denoted by r and lower tail denoted by l. m and s are Normal mean and standard deviation.

Figure 5.7.1 *Upper Tail*

Figure 5.7.2 *Lower Tail*

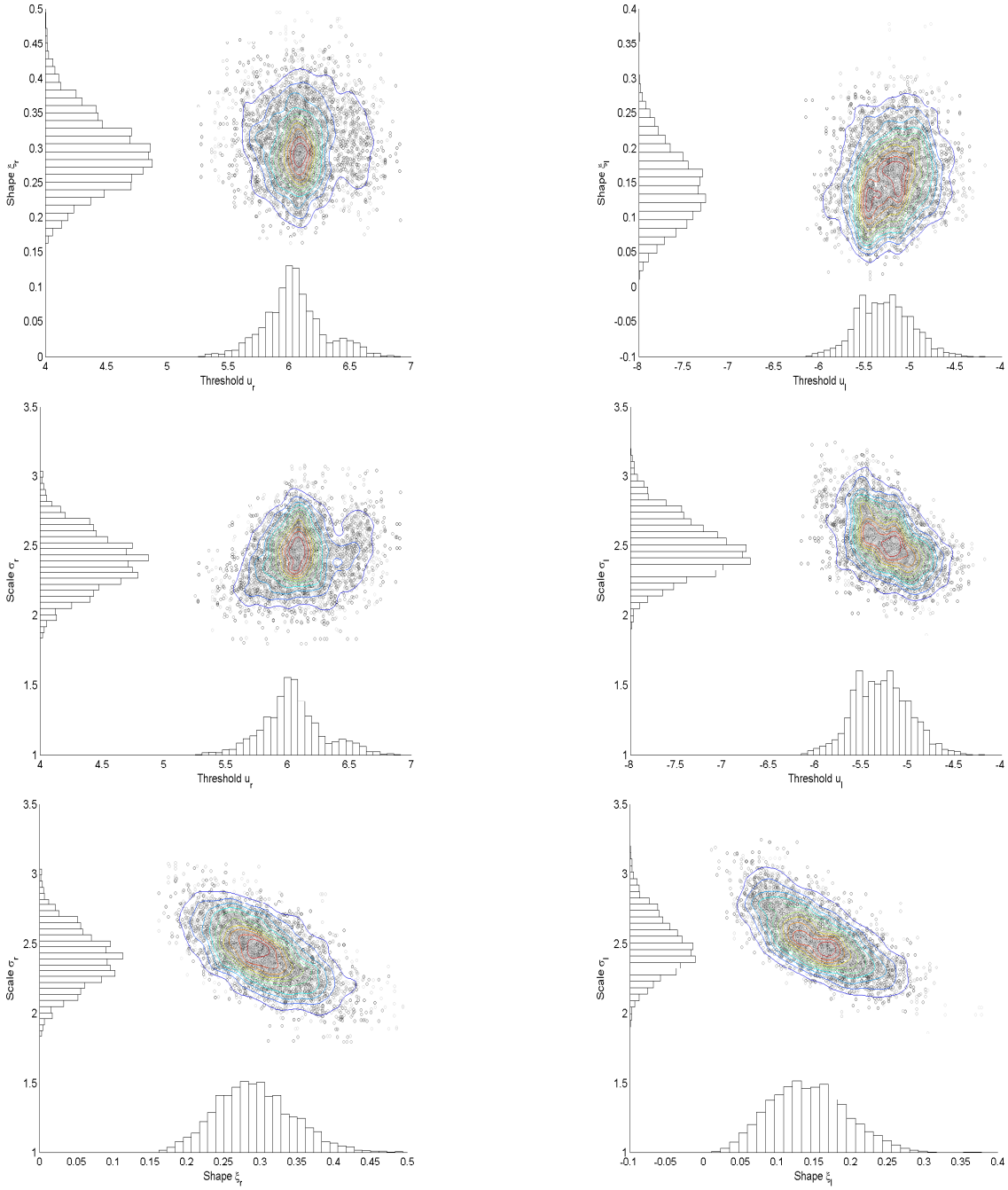

Figure 5.7: Pairwise contours of the posterior density of the GPD related parameters for the lower tail $(\hat{u_l}, \hat{\sigma_l}, \hat{\xi_l})$ and upper tail $(\hat{u_r}, \hat{\sigma_r}, \hat{\xi_r})$. the histogram of each posterior distribution is also shown.

bution becomes more skewed, representing the asymmetry in the information available for estimation.

In order to see the impacts of threshold choice on the parameter estimation and inference afterwards, the standard GPD with pre-fixed threshold method is conducted on the simulation sample to get the shape and scale estimation for both tails. Table 5.2 gives

Figure 5.8.1 *Upper Tail*

Figure 5.8.2 *Lower Tail*

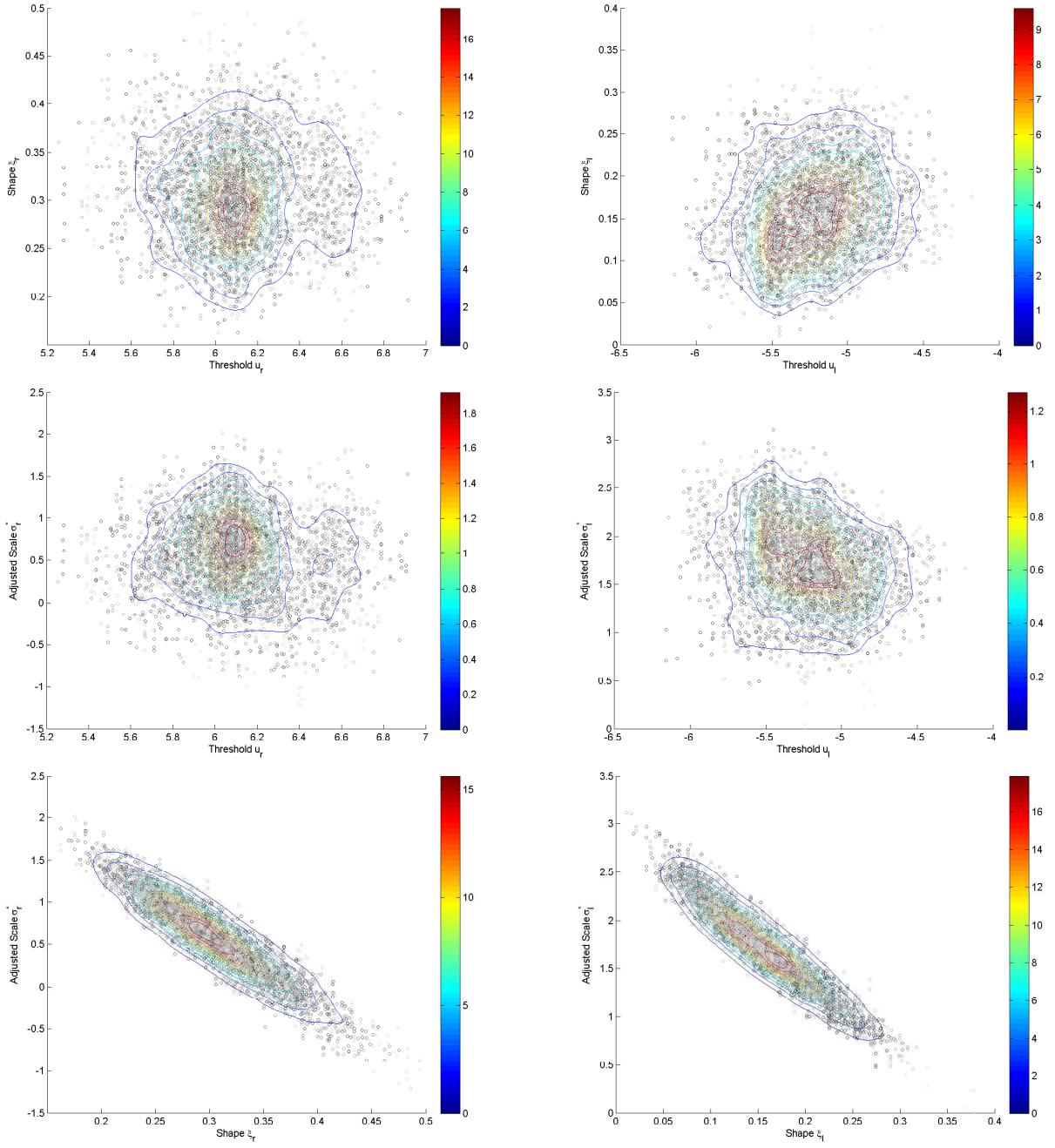

Figure 5.8: Adjusted pairwise contours of the posterior density of the GPD related parameters for the lower tail $(\hat{u_l}, \hat{\sigma_l^*}, \hat{\xi_l})$ and upper tail $(\hat{u_r}, \hat{\sigma_r^*}, \hat{\xi_r})$.

the GPD parameter estimators based on standard GPD with different threshold values. In application, the decision is hard to make. On the other hand, the mixture model find a reasonable threshold value and also obtain the uncertainty interval for it.

The quantile estimate based on standard GPD with fixed threshold value and the GNG are reported in Table 5.3 as an example. The variation of the high quantiles (such as 0.01% and 0.99%) is larger than those quantiles closer to the mode of the distribution

Table 5.2: Parameter estimates for GNG and standard GPD.*Note:* POE represent the proportion above the upper threshold or the proportion below the lower threshold. The threshold value used for each fixed GPD model is the quantile of $1 - POE$ for upper threshold and POE for lower threshold.

Model	**POE**	u	ξ	**CI of** ξ		σ	**CI of** σ	
Upper Tail								
Fixed GPD	0.15	4.3222	0.2837	0.1989	0.3686	2.2524	2.0195	2.5121
	0.12	4.8860	0.3088	0.2114	0.4061	2.2902	2.0240	2.5915
	0.08	5.9060	0.3554	0.2291	0.4816	2.4139	2.0652	2.8214
	0.07	6.2823	0.3777	0.2391	0.5164	2.4292	2.0510	2.8771
GNG		*6.0986*	*0.3015*	*0.2038*	*0.4055*	*2.4318*	*2.0456*	*2.8273*
True Pars		**6.0000**	**0.3000**			**2.2000**		
Lower Tail								
Fixed GPD	0.15	-4.3996	0.1828	0.0940	0.2717	2.3437	2.0926	2.6249
	0.12	-4.9401	0.1858	0.0841	0.2874	2.4280	2.1352	2.7610
	0.08	-5.9140	0.1436	0.0234	0.2638	2.7942	2.3943	3.2608
	0.07	-6.2882	0.1409	0.0114	0.2705	2.8599	2.4226	3.3761
GNG		*-5.2267*	*0.1532*	*0.0583*	*0.2572*	*2.5251*	*2.1555*	*2.9343*
True Pars		**-5.0000**	**0.2000**			**2.5000**		

as we would expect. The difference among the quantile estimators for the different thresholds in the fixed threshold approach is substantial. The difference is larger for the heavier tailed distribution as you would expect. Again the estimators from the GNG model is close to the true quantile values and the differences are due to natural sample variability.

Table 5.3: Quantile estimators with different threshold choices. *Note:* POE represent the proportion above the upper threshold or the proportion below the lower threshold. The threshold value used for each Fixed GPD model is the quantile of $1 - POE$ for upper threshold and POE for lower threshold.

Model	**Quantile Estimators**								
Fixed GPD	**POE**	**0.0001**	**0.001**	**0.01**	**0.1**	**0.9**	**0.99**	**0.999**	**0.9999**
	0.15	-40.39	-23.62	-12.61	-5.39	5.29	13.50	29.28	59.61
	0.12	-40.66	-23.68	-12.61	-5.39	5.32	13.44	29.99	63.68
	0.08	-37.27	-22.96	-12.69	Nav	Nav	13.34	31.35	72.18
	0.07	-37.08	-22.93	-12.69	Nav	Nav	13.26	31.86	76.22
GNG		-34.04	-22.92	-12.48	-5.38	5.37	12.76	28.33	51.98
Sample Quans		-28.64	-23.10	-12.76	-5.37	5.39	13.20	28.79	156.97
True Quans		-43.84	-24.90	-12.94	-5.40	5.38	12.17	25.61	52.43

5.4.2 Multiple Simulation Results

Six parameter sets from the GNG model have been chose to simulate data from populations which different tail behaviors and sample size. Table 5.4 reports the true parameters and properties of estimated parameters of these samples. Parameter set *1* represents a population with both sides of type I Pareto tails with $\xi = 0$. Parameter set *2* represents a population with both sides of type II Pareto tails (or heavy tailed distribution) with $\xi > 0$. Parameter set *3* represents a population with both sides of type III Pareto tails (or upper bounded tails) with $\xi < 0$. Parameter set *4* to *6* represent the different combinations of the three type tails. All parameter sets are not symmetric distributed in tails. The parameter values chosen have been chosen to ensure the populations have a smooth density at the thresholds, as this is the most physically reasonable. There are three sample size (1000, 3000 and 5000) used for each parameter set. For each parameter set with a particular size, 100 samples are simulated. The reported results in the table is the mean value over these 100 samples. For each sample estimation, the MCMC chain choose the length of 10000 and the first half is discarded. The estimated parameters are calculated as the mean of 95% highest posterior density interval (HPD interval) of 2nd half of the chain, and the credible interval (CI) of estimators are the boundary of the associated HPD interval.

All the estimates for these simulated data are close to the real parameters and the 95% credible intervals for all of them contain the true parameter values. Table 5.4 also reports the square root of mean square error (RMSE) of the estimated parameters. It is clear that the MSE decreases as the sample size increases for all the parameters which means a larger variance and wider CI associated with smaller samples. In particular, the efficiency increases faster for the heavier tail compare to the short tail as the size increase since the sparsity of extremes for heavy tail distributions than others. Generally, the MSE for the threshold has the largest value and the MSE for the normal related parameters has the smallest value as expected, which indicates that the threshold is the most difficult one to estimate in the mixture model and the normal related parameters are easiest to estimate since more information contained in the data for the bulk distribution. Further, we also know that a range of different thresholds can provide suitable model fits. The traditional graphical diagnostics explicitly show that a range of thresholds are appropriate.

This simulation study shows the reliability in the model estimation. Even for the smallest sample size 1000, the model can still manage to find a close value to the true parameters and describe both tails accurately. The simulation also shows the flexibility of the model in capturing all sorts of tail behaviors.

Table 5.4: Summary of properties of the Bayesian estimates of the GNG model parameters, for a range of different parameters sets (tail behaviours). There are 100 simulated datasets for parameter set. The true parameters along with the mean and root mean square error (RMSE) of the point estimates across the 100 sample estimations. The point estimates for each sample are the mean of the posterior within the 95% highest posterior density.

1. I-N-I	**True**	**size=1000**		**size=3000**		**size=5000**	
Param	**Value**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**
m	0.00	0.0081	0.0416	0.0019	0.0228	-0.0004	0.0182
s	2.00	2.0087	0.0334	2.0055	0.0181	2.0042	0.0132
u_r	2.30	2.2524	0.1839	2.2774	0.1940	2.3084	0.1689
ξ_r	0.00	-0.0339	0.0981	-0.0156	0.0538	-0.0064	0.0426
σ_r	1.20	1.2401	0.1552	1.2189	0.0844	1.2170	0.0799
u_l	-2.50	-2.2895	0.2706	-2.3570	0.2568	-2.3442	0.2564
ξ_l	0.00	-0.0299	0.0877	-0.0065	0.0562	-0.0130	0.0444
σ_l	1.15	1.1915	0.1699	1.1577	0.0948	1.1776	0.0817
2. II-N-II	**True**	**size=1000**		**size=3000**		**size=5000**	
Param	**Value**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**
m	0.00	0.0045	0.0342	-0.0005	0.0201	0.0000	0.0150
s	2.00	2.0139	0.0304	2.0055	0.0158	2.0070	0.0138
u_r	2.00	2.0953	0.2418	2.1138	0.2236	2.0577	0.2150
ξ_r	0.20	0.1575	0.0857	0.1868	0.0614	0.1851	0.0518
σ_r	1.30	1.3334	0.1586	1.3354	0.1205	1.3294	0.0913
u_l	-1.80	-1.9485	0.2438	-1.8757	0.1796	-1.8498	0.1635
ξ_l	0.30	0.2715	0.0861	0.2793	0.0605	0.2924	0.0438
σ_l	1.40	1.4496	0.1950	1.4446	0.1211	1.4249	0.0902
3. III-N-III	**True**	**size=1000**		**size=3000**		**size=5000**	
Param	**Value**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**
m	0.00	0.0031	0.0493	-0.0004	0.0200	-0.0023	0.0194
s	2.00	2.0046	0.0334	2.0014	0.0186	2.0025	0.0146
u_r	2.30	2.3811	0.2951	2.3721	0.2860	2.3660	0.3185
ξ_r	-0.30	-0.2865	0.0898	-0.2947	0.0592	-0.2971	0.0469
σ_r	1.20	1.1721	0.1951	1.1704	0.1145	1.1770	0.0839
u_l	-2.50	-2.4820	0.3477	-2.4782	0.3267	-2.4580	0.3075
ξ_l	-0.20	-0.1722	0.0827	-0.2078	0.0614	-0.1941	0.0451
σ_l	1.20	1.1671	0.1452	1.2259	0.1277	1.2097	0.1057

Table 5.5: Summary of properties of the Bayesian estimates of the GNG model parameters, for a range of different parameters sets (tail behaviours). There are 100 simulated datasets for parameter set. The true parameters along with the mean and root mean square error (RMSE) of the point estimates across the 100 sample estimations. The point estimates for each sample are the mean of the posterior within the 95% highest posterior density.

4. III-N-II	**True**	**size=1000**		**size=3000**		**size=5000**	
Param	**Value**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**
m	0.00	0.0108	0.0405	0.0004	0.0251	-0.0005	0.0192
s	2.00	2.0088	0.0305	2.0038	0.0167	2.0023	0.0139
u_r	2.50	2.3155	0.1886	2.3534	0.2267	2.4185	0.1978
ξ_r	0.20	0.1747	0.0865	0.1909	0.0457	0.1895	0.0357
σ_r	1.15	1.1291	0.1536	1.1404	0.1204	1.1482	0.0948
u_l	-2.60	-2.4365	0.1637	-2.5101	0.2035	-2.5190	0.2162
ξ_l	-0.15	-0.1476	0.0934	-0.1439	0.0561	-0.1413	0.0390
σ_l	1.10	1.1226	0.1714	1.1146	0.1100	1.1093	0.0841
5. III-N-I	**True**	**size=1000**		**size=3000**		**size=5000**	
Param	**Value**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**
m	0.00	-0.0123	0.0418	-0.0035	0.0214	-0.0031	0.0190
s	2.00	2.0112	0.0312	2.0022	0.0204	2.0035	0.0132
u_r	2.70	2.4599	0.2791	2.4837	0.2421	2.4483	0.2117
ξ_r	0.00	-0.0181	0.1111	-0.0114	0.0632	0.0005	0.0500
σ_r	1.30	1.2949	0.2716	1.3013	0.0941	1.2777	0.0744
u_l	-2.80	-2.4900	0.2225	-2.5516	0.2125	-2.5718	0.2522
ξ_l	-0.10	-0.1078	0.0796	-0.1132	0.0547	-0.1077	0.0455
σ_l	1.00	1.0514	0.1639	1.0638	0.1079	1.0562	0.1041
6. II-N-I	**True**	**size=1000**		**size=3000**		**size=5000**	
Param	**Value**	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**
m	0.00	0.0073	0.0468	0.0060	0.0241	0.0026	0.0172
s	2.00	2.0121	0.0342	2.0070	0.0190	2.0086	0.0166
u_r	2.30	2.2478	0.2066	2.2670	0.1923	2.2755	0.1616
ξ_r	0.15	0.1387	0.1029	0.1447	0.0576	0.1500	0.0440
σ_r	1.20	1.1854	0.1781	1.1955	0.1034	1.1971	0.0703
u_l	-2.50	-2.3282	0.2853	-2.3551	0.2367	-2.3419	0.2391
ξ_l	0.00	-0.0219	0.1001	-0.0083	0.0512	-0.0069	0.0403
σ_l	1.20	1.2357	0.1961	1.2154	0.0980	1.1997	0.0727

5.5 Simulation Study II - Performance For General Distributions

The previous simulation results show the performance of the estimation method for the GNG mixture model. However, in real applications of this approach the population will be approximated by the GNG model. Therefore in this section, a sensitivity analysis is conducted by applying the GNG model to various population distributions, including both symmetric and asymmetric distributions, to show how the model performs as an approximation in this case.

The distributions chosen for this simulation study are symmetric distributions: Normal (type I tails), t (type II tails) and symmetric Beta (type III tails); and asymmetric distributions: Gumbel (type III and type I tails), inverse gamma (type III and type II tails) and Weibull (type III and type III tails).

For each distribution, we simulated 100 datasets with sample sizes 1000, 3000 and 5000 as before. The estimation results are shown as Table 5.6. The shape parameters for all six different distributions are as expected sign as their true tail behaviors (zero shape for type I, positive shape for type II and negative shape for type III). Also the biases of the estimated shape parameter from the true shape decrease as the sample size increases, e.g. the shape is close to zero for the normal distribution. There are generally negative biases of the shape estimators due to lack of the asymptotic convergence of the GPD. The calculated variance of the estimate for the multiple sample simulations shows the asymptotic convergence for the normal tail is slower than other distributions and uncertainty of the estimators increases as the sample size decrease and asymmetry of the bulk data increase.

The variability of the estimate across distributions and sample sizes is reported in Table 5.7. For symmetric distributions, the standard deviations reduce when the sample size increase for both normal parameters and GPD parameters. However, the variation does not reduce for asymmetric distributions as fast as the symmetric distributions since the GNG model assume the symmetric feature around the mode, which makes threshold estimation more problematic. In particular, the variability in estimating the threshold for Weibull increases as the sample size increases and consequently an increase for the uncertainty of GPD scale (The fitted density plot for these distributions are shown in Appendix A.1). These results indicates that the choice of bulk distribution can affect the fitting of the tail distribution and it should be carefully examined according to applications. Alternatively, a less restricted distribution can be used for bulk data or a non-parametric method can increase the flexibility of the model. The benefit of the

normal in the research is it's simplicity and it being appropriate for the financial applications concerned later on, due to their inherent unimodal, approximately symmetric and quadratic shape around the mode.

Table 5.6: Estimation Results of the GNG model on the General Distributions

Distributions	**Sample Size**	**Mean of Parameter Estimations**							
		m	s	u_r	ξ_r	σ_r	u_l	ξ_l	σ_l
Normal	1000	0.0138	2.0263	2.5908	-0.0131	1.0252	-2.6114	-0.0163	1.0356
	3000	0.0047	2.0030	3.2332	-0.0170	0.9300	-3.3893	0.0121	0.8619
	5000	0.0003	2.0051	3.4514	-0.0112	0.8926	-3.5557	-0.0054	0.8630
T	1000	0.0034	1.1719	0.9005	0.1671	0.8872	-0.9172	0.1863	0.8718
	3000	0.0010	1.1476	0.8859	0.1983	0.8632	-0.8659	0.2057	0.8637
	5000	-0.0015	1.1439	0.8465	0.1869	0.8756	-0.8440	0.1859	0.8774
Beta	1000	0.4998	0.1249	0.6611	-0.2392	0.0654	0.3389	-0.2197	0.0631
	3000	0.5003	0.1248	0.6601	-0.2190	0.0625	0.3388	-0.2262	0.0631
	5000	0.5000	0.1244	0.6604	-0.2282	0.0631	0.3399	-0.2272	0.0634
Gumbel	1000	0.3949	1.1759	1.1752	-0.0606	1.1604	-0.2804	-0.2955	0.6293
	3000	0.3843	1.0884	0.6812	-0.0696	1.2189	-0.3882	-0.2856	0.5801
	5000	0.3691	1.0532	0.5293	-0.0702	1.2359	-0.4233	-0.2902	0.5641
Inverse Gamma	1000	1.0918	0.6177	1.7621	0.1329	0.6209	0.9083	-0.4573	0.2856
	3000	1.0868	0.4890	1.3358	0.1454	0.5772	0.8292	-0.4224	0.2357
	5000	1.0785	0.4515	1.1974	0.1307	0.5709	0.8170	-0.4094	0.2246
Weibull	1000	2.1604	5.6826	9.7939	-0.1130	4.2164	-3.8944	-0.1732	2.1970
	3000	2.1192	5.7106	9.3348	-0.1189	4.2993	-3.1004	-0.2286	2.5488
	5000	2.0843	5.7169	8.8573	-0.1247	4.3896	-2.6406	-0.2405	2.7399

Table 5.8 shows the performance of the GNG model to approximate various quantiles of the population distributions. The true and GNG estimated quantiles are shown in the upper part of the table. It is clear that all the GNG quantile estimates for all six distributions are close to the true values.

The RMSE of the posterior quantile estimates for the six distributions are reported in the lower part of Table 5.9. For comparison purposes, the RMSE of the quantile estimators using maximum likelihood (ML) estimation when using the correct population distributions in the model. The RMSE for the correct model using ML estimation is considered a gold standard, to compare the performance of the approximate GNG model. Although the ML confidence intervals and Bayesian credible intervals are not formally comparable, as very diffuse priors have been used when estimating the parameters for the GNG model, the two sets of intervals are practically comparable. The ML quantiles are estimated under the correct model and therefore have a smaller RMSE. However, you will notice that ML estimation under the correct model is at most twice as efficient for estimating the various quantiles compared to the GNG model. Overall the GNG approximation is only slightly less efficient for most population quantiles, compared to using the correct population distribution model.

Table 5.7: Variation of the GNG model fitting on the General Distributions

Distributions	**Sample Size**	**Standard Deviation of Parameter Estimations**							
		m	s	u_r	ξ_r	σ_r	u_l	ξ_l	σ_l
Normal	1000	0.0696	0.0685	0.8094	0.1265	0.2939	0.7984	0.1155	0.2792
	3000	0.0384	0.0272	0.7792	0.1262	0.2325	0.7912	0.1168	0.2298
	5000	0.0284	0.0228	0.8862	0.1298	0.2401	0.7700	0.1225	0.2085
T	1000	0.0458	0.0593	0.2605	0.0787	0.0836	0.2530	0.0820	0.0781
	3000	0.0243	0.0333	0.1555	0.0517	0.0455	0.1734	0.0525	0.0411
	5000	0.0192	0.0214	0.1337	0.0435	0.0396	0.1415	0.0446	0.0382
Beta	1000	0.0042	0.0032	0.0045	0.1056	0.0086	0.0055	0.1045	0.0088
	3000	0.0020	0.0019	0.0052	0.0592	0.0056	0.0047	0.0627	0.0060
	5000	0.0020	0.0014	0.0049	0.0429	0.0040	0.0054	0.0445	0.0041
Gumbel	1000	0.0486	0.1090	0.3935	0.0649	0.1244	0.3045	0.1117	0.1761
	3000	0.0320	0.0876	0.2870	0.0365	0.0819	0.2612	0.0725	0.1342
	5000	0.0246	0.0634	0.1868	0.0254	0.0642	0.2072	0.0482	0.1007
Inverse Gamma	1000	0.0300	0.0666	0.1370	0.0772	0.0747	0.0933	0.0736	0.0713
	3000	0.0174	0.0728	0.1717	0.0445	0.0279	0.0909	0.0677	0.0623
	5000	0.0129	0.0466	0.1021	0.0306	0.0187	0.0649	0.0445	0.0420
Weibull	1000	0.1921	0.1685	0.4081	0.0689	0.5191	0.5153	0.0684	0.2484
	3000	0.1141	0.1409	0.5543	0.0439	0.3285	0.6915	0.0447	0.3047
	5000	0.0957	0.1437	0.6225	0.0363	0.2936	0.7895	0.0424	0.3526

The uncertainty of the quantiles is higher for heavier tails (t and upper tail of inverse gamma) compared to the short tails (e.g. beta and lower tail of Gumbel). The normal distribution has a slightly higher RMSE than the others due to known slow asymptotic convergence of the normal tail to the GPD limit (see Beirlant *et al*, 2004). As expected the RMSE increases as the quantile is located further out into the tail of the distribution. The differences between posterior predictive quantiles and the ML quantile estimates are smaller for the heavy tails compare to the short or exponential tails. This result indicates that the GNG model is preferred when describing the tail behavior for the heavy tail applications, which is the typical case found in finance and economic applications.

It is clear from Table 5.9 that population distributions with highly asymmetric modes result in higher uncertainty, since the normal distribution is used for the bulk of distribution in the GNG model. However, the mixture model can still return reasonable extreme quantile estimates. This results shows that the GNG model is generally applicable as an approximation to a wide range of population distributions.

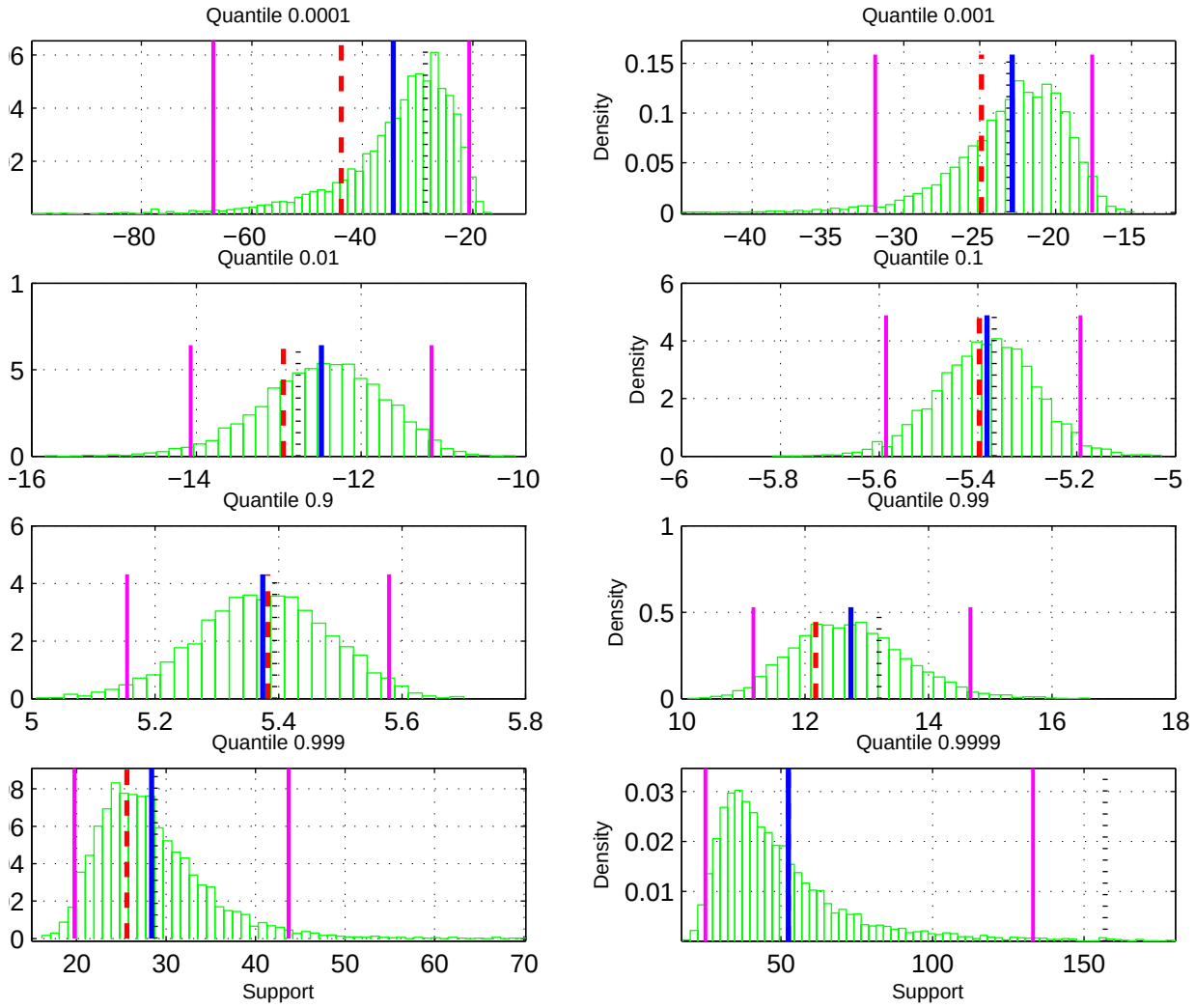

Figure 5.9: Histogram of the posterior predictive quantiles for unconditional tail probabilities 0.01%, 0.1%, 1% and 10%. The sample quantile is shown by the dotted line and the true quantile by the interior dashed line. The posterior predictive quantile (PPQ) shown by the solid line is the mean of posterior predictive quantiles within the 95% credible intervals, shown by the exterior solid lines.

Table 5.8: Comparison of quantiles estimates for GNG model applied to various general population distributions (fitted using Bayesian inference) and true model distribution (fitted using maximum likelihood estimation). The mean of the posterior predictive quantiles (PPQ) within the 95% highest posterior density have been used as point estimates for each sample. The mean and RMSE of the point estimates are obtained from 100 simulated datasets.

Distributions	**Sample Size**	**Posterior Predictive Quantiles**											
		1%		**2%**		**5%**		**95%**		**98%**		**99%**	
		TRUE	**PPQ**	**TRUE**	**PPQ**	**TRUE**	**PPQ**	**TRUE**	**PPQ**	**TRUE**	**PPQ**	**TRUE**	**PPQ**
	1000	-4.653	-4.709	-4.107	-4.130	-3.290	-3.290	3.290	3.301	4.107	4.132	4.653	4.704
Normal	3000	-4.653	-4.648	-4.107	-4.098	-3.290	-3.278	3.290	3.293	4.107	4.118	4.653	4.672
$(\mu = 0, \sigma = 2)$	5000	-4.653	-4.663	-4.107	-4.112	-3.290	-3.287	3.290	3.294	4.107	4.118	4.653	4.667
	1000	-4.541	-4.518	-3.482	-3.495	-2.353	-2.351	2.353	2.370	3.482	3.497	4.541	4.485
t	3000	-4.541	-4.625	-3.482	-3.559	-2.353	-2.375	2.353	2.363	3.482	3.525	4.541	4.563
$(v = 3)$	5000	-4.541	-4.576	-3.482	-3.550	-2.353	-2.387	2.353	2.381	3.482	3.543	4.541	4.569
	1000	0.229	0.225	0.256	0.254	0.300	0.299	0.700	0.702	0.744	0.747	0.771	0.776
Beta	3000	0.229	0.227	0.256	0.255	0.300	0.299	0.700	0.700	0.744	0.745	0.771	0.773
$(\alpha = 8, \beta = 8)$	5000	0.229	0.227	0.256	0.255	0.300	0.300	0.700	0.700	0.744	0.745	0.771	0.773
	1000	-1.527	-1.562	-1.364	-1.393	-1.097	-1.109	2.970	2.959	3.902	3.891	4.600	4.568
Gumbel	3000	-1.527	-1.568	-1.364	-1.395	-1.097	-1.107	2.970	3.015	3.902	3.949	4.600	4.619
$(\sigma = 1)$	5000	-1.527	-1.567	-1.364	-1.394	-1.097	-1.106	2.970	3.018	3.902	3.958	4.600	4.631
Inverse	1000	0.431	0.410	0.473	0.453	0.546	0.536	2.538	2.451	3.269	3.148	3.909	3.740
Gamma	3000	0.431	0.414	0.473	0.458	0.546	0.540	2.538	2.542	3.269	3.281	3.909	3.913
$(\alpha = 5, \beta = 5)$	5000	0.431	0.413	0.473	0.458	0.546	0.542	2.538	2.557	3.269	3.287	3.909	3.903
	1000	-8.250	-8.577	-7.307	-7.550	-5.798	-6.007	12.849	12.175	16.154	15.588	18.436	17.953
Weibull	3000	-8.250	-8.473	-7.307	-7.490	-5.798	-5.932	12.849	12.324	16.154	15.738	18.436	18.090
$(\lambda = 5, k = 0.1)$	5000	-8.250	-8.483	-7.307	-7.492	-5.798	-5.903	12.849	12.440	16.154	15.846	18.436	18.179

Table 5.9: Comparison of quantiles estimates for GNG model applied to various general population distributions (fitted using Bayesian inference) and true model distribution (fitted using maximum likelihood estimation). The mean of the posterior predictive quantiles (PPQ) within the 95% highest posterior density have been used as point estimates for each sample. The mean and RMSE of the point estimates are obtained from 100 simulated datasets.

Distributions	**Sample Size**	**RMSE of Quantile Estimates**											
		1%		**2%**		**5%**		**95%**		**98%**		**99%**	
		ML	**PPQ**	**ML**	**PPQ**	**ML**	**PPQ**	**ML**	**PPQ**	**ML**	**PPQ**	**ML**	**PPQ**
	1000	0.130	0.216	0.119	0.168	0.103	0.120	0.107	0.115	0.123	0.156	0.135	0.210
Normal	3000	0.068	0.114	0.063	0.083	0.055	0.061	0.052	0.060	0.059	0.084	0.064	0.104
$(\mu = 0, \sigma = 2)$	5000	0.056	0.076	0.051	0.062	0.045	0.053	0.044	0.049	0.051	0.069	0.056	0.083
	1000	0.372	0.390	0.221	0.236	0.098	0.123	0.098	0.132	0.221	0.241	0.372	0.383
t	3000	0.208	0.263	0.123	0.168	0.055	0.082	0.055	0.075	0.123	0.136	0.208	0.209
$(v = 3)$	5000	0.162	0.184	0.097	0.131	0.043	0.074	0.043	0.076	0.097	0.138	0.162	0.199
	1000	0.006	0.009	0.006	0.008	0.005	0.006	0.005	0.006	0.006	0.008	0.006	0.010
Beta	3000	0.004	0.006	0.004	0.005	0.003	0.004	0.003	0.003	0.003	0.005	0.003	0.006
$(\alpha = 8, \beta = 8)$	5000	0.003	0.004	0.003	0.004	0.002	0.003	0.002	0.003	0.002	0.003	0.003	0.004
	1000	0.053	0.071	0.048	0.060	0.041	0.044	0.105	0.125	0.161	0.167	0.216	0.219
Gumbel	3000	0.028	0.055	0.025	0.044	0.021	0.025	0.049	0.074	0.081	0.098	0.113	0.126
$(\sigma = 1)$	5000	0.023	0.050	0.021	0.040	0.018	0.024	0.046	0.072	0.073	0.094	0.099	0.107
Inverse	1000	0.010	0.027	0.010	0.024	0.010	0.016	0.080	0.131	0.128	0.198	0.176	0.275
Gamma	3000	0.005	0.020	0.005	0.017	0.005	0.010	0.044	0.048	0.069	0.088	0.094	0.145
$(\alpha = 5, \beta = 5)$	5000	0.004	0.019	0.004	0.016	0.005	0.007	0.031	0.040	0.050	0.070	0.069	0.111
	1000	0.269	0.450	0.241	0.355	0.206	0.304	0.359	0.799	0.491	0.796	0.619	0.805
Weibull	3000	0.149	0.284	0.133	0.231	0.113	0.181	0.236	0.588	0.345	0.551	0.447	0.564
$(\lambda = 5, k = 0.1)$	5000	0.132	0.285	0.117	0.227	0.098	0.150	0.164	0.467	0.233	0.426	0.300	0.427

5.6 Robust Threshold Estimation

As previously discussed, there are many techniques which can be used to estimate the threshold (see for example Coles (2001) and Beirlant, Goegebeur, Segers, and Teugels (2004)). Typically these techniques have been difficult to automate, often requiring manual intervention requiring subjective judgement. Dupuis (1998) suggests a robust threshold selection method which examines the weights applied to the extremes to assess the validity of the GPD under a range of proposed thresholds. Specifically, at a given threshold, estimate the GPD parameters for the excess above the threshold using the optimal bias robust estimator (OBRE) which fits the bulk of the data well. The meantime, OBRE assigns a weight to each data point and low weight indicates the data do not fit the model by the robust estimates. Therefore, those weight with the extremes can be used to address the threshold choices as a too low threshold will cause a lack of fit of GPD and many extremes assigned low weights. The principle of threshold choice should chose the value as low as possible to maximize the tail information and satisfy the asymptotic property. We implement this robust threshold estimation method to check our threshold estimation method and the resultant tail fits from the proposed mixture model.

Dupuis' method of robust threshold choice is applied to check the threshold estimations and the tail fitting from the proposed mixture model with the robust estimation of GPD. We apply the robust GPD estimation, using the algorithm as Hample, Ronchetti, Rousseeuw, and Stahel (1986) and Victoria-Feser and Ronchetti (1994), on various simulated data and the thresholds suggested by the robust estimation are consistent from those suggested by the proposed mixture model. We do not report the full results for brevity and only show one of the simulation example to illustrate. Table 5.10 lists the true parameter values used for the simulation sample with the parameters estimated through OBRE procedure. Table 5.11 and Table 5.12 lists the weights of the 26 largest data points for the upper tail and the weights of the 26 smallest data points for the lower tail. The tables suggest a upper threshold value of 1.6 and lower threshold of -1.8, which are consistent with the mixture model estimates. Although these weights of extremes are useful in threshold selections, the method is still only semi-automated and certain subjective decision still needed. Further, the Dupuis' method is a fixed-threshold approach and therefore can not capture the uncertainty of the threshold estimate. The method is not a convenient way in forecasting and inference since the supervision is required in the threshold selection. On the other hand, the proposed mixture model overcome these disadvantages with the automated estimate of the threshold.

Table 5.10: Comparison of the estimated threshold and GPD parameters using the GNG mixture model and Robust estimation procedure of Dupuis (1998).

	Upper Tail			Lower Tail		
	u_r	ξ_r	σ_r	u_l	ξ_l	σ_l
True Parameters	2.00	0.20	1.30	-1.80	0.30	1.40
Mixture Estimates	1.64	0.17	1.39	-1.77	0.29	1.30
Robust Estimates	1.60	0.17	1.40	-1.80	0.26	1.31

5.7 Asymmetry Analysis

In applications, a parsimonious model is always preferred. With the asymmetry assumption of tails in the proposed GNG model, the additional three parameters are required. This induces the question as to whether the model is over-fitted and the identification of the asymmetry feature. The asymmetry feature of financial data has been well documented (Glosten, Jagannathan, and Runkle 1992; Bollen and Whaley 2004). Therefore, it is worthwhile conducting a statistical analysis to the complex model (asymmetric model) and simpler model (symmetric model) through Bayesian inference. The complex model has the advantage in fitting the data but disadvantage of being more degrees of freedom used up. The key point of comparison is therefore to identify whether the improvements in model is large enough to justify the additional degree of freedom.

In this research, we prefer the approach that measure the distance of the data to each of the approximate models. In Bayesian inference, this is equivalent to comparing two posterior distributions $p(\theta|x)$ and $p(\theta, \varphi|x)$, where θ is the parameter of the simpler model and φ is the extra parameters of the complex model, along with their predictive distributions for replicated data. We compare the overall model fitting in measuring estimated predictive discrepancy, using prediction error–deviance $D(x, \theta)$, which is defined as:$D(x|\theta) = -2\log(p(x|\theta))$. The expected deviance does not depend on θ. The discrepancy between data and model depends in general on θ as well as x. The deviance at $\hat{\theta}$ is defined as: $D_{\hat{\theta}}(x) = D(x, \hat{\theta}(x))$, where $\hat{\theta}$ is the mean of the posterior distribution. The posterior mean deviance is defined as $D_{avg}(x) = E(D(x), x)$. It can be calculated as the average of discrepancy over the posterior distributions as: $\hat{D}_{avg}(x) = \frac{1}{k}\sum_{k=1}^{k} D(x|\theta_k)$, here K is the length of posterior simulations. The difference between the posterior mean deviance and the deviance at $\hat{\theta}$, $D_D = D_{avg}(x) - D_{\hat{\theta}}(x)$, represents the effect of model fitting as the measure of the effective number of unconstrained parameters of a Bayesian model. It equals the number of 'unconstrained' parameters in the model. The deviance information criterion (DIC) is defined as $DIC = D_{avg}(x) + D_D$.

The data firstly used for asymmetry analysis is a simulated data from the mixture

Table 5.11: Robust estimate of GPD of the upper tail of simulated sample. *Note:* The estimates of GPD parameters and weight of extremes of upper tail of simulated sample are produced by robust estimation using a range of pre-fixed threshold values of classical GPD. The number of excess is the number of observations above the threshold given. The associated weights from the robust estimation for the largest 26 data points are listed in the table. The return level is the quantile at 1-p (98%) in the table for different threshold choices.

Robust Estimates and Weights of Upper Tail							
u	1.4	*1.6*	1.8	2	2.2	2.4	2.6
Numb of Excess	735	*629*	545	476	420	370	330
Robust ξ	0.188	*0.174*	0.176	0.182	0.206	0.233	0.296
Robust σ	1.334	*1.400*	1.437	1.463	1.456	1.451	1.384
Return level (p=0.02)	9.106	*9.450*	9.890	10.352	10.955	11.665	12.810

Rank	**weight**	**weight**	**weight**	**weight**	**weight**	**weight**	**weight**
26	0.9897	*1*	1	1	1	1	1
25	0.9817	*1*	1	1	1	1	1
24	0.9435	*1*	1	1	1	1	1
23	0.9378	*1*	1	1	1	1	1
22	0.9338	*1*	1	1	1	1	1
21	0.9294	*1*	1	1	1	1	1
20	0.8148	*0.9108*	1	1	1	1	1
19	0.7940	*0.8859*	1	1	1	1	1
18	0.7512	*0.8347*	0.9435	1	1	1	1
17	0.7437	*0.8258*	0.9331	1	1	1	1
16	0.6874	*0.7588*	0.8542	0.9602	1	1	1
15	0.6639	*0.7308*	0.8213	0.9220	1	1	1
14	0.6518	*0.7165*	0.8044	0.9024	1	1	1
13	0.5707	*0.6207*	0.6922	0.7720	0.8734	0.9943	1
12	0.5638	*0.6126*	0.6827	0.7610	0.8609	0.9799	1
11	0.5438	*0.5891*	0.6553	0.7292	0.8246	0.9385	1
10	0.4844	*0.5200*	0.5748	0.6363	0.7187	0.8172	0.9498
9	0.4714	*0.5049*	0.5574	0.6163	0.6958	0.7910	0.9209
8	0.4622	*0.4942*	0.5450	0.6020	0.6796	0.7724	0.9003
7	0.4536	*0.4843*	0.5335	0.5888	0.6646	0.7553	0.8813
6	0.4047	*0.4283*	0.4691	0.5151	0.5808	0.6595	0.7748
5	0.3484	*0.3644*	0.3962	0.4324	0.4871	0.5527	0.6554
4	0.3437	*0.3591*	0.3902	0.4255	0.4793	0.5439	0.6455
3	0.2917	*0.3010*	0.3246	0.3519	0.3963	0.4496	0.5393
2	0.2702	*0.2773*	0.2981	0.3222	0.3630	0.4119	0.4965
1	0.2568	*0.2626*	0.2817	0.3040	0.3425	0.3888	0.4703

Table 5.12: Robust estimate of GPD of the lower tail of simulated sample. *Note:* The estimates of GPD parameters and weight of extremes of lower tail of simulated sample are produced by robust estimation using a range of pre-fixed threshold values of classical GPD. The number of excess is the number of observations below the threshold given. The associated weights from the robust estimation for the smallest 26 data points are listed in the table. The return level is the quantile at p (2%) in the table for different threshold choices.

Robust Estimates and Weights of Lower Tail							
Threshold	-1.4	-1.6	*-1.8*	-2	-2.2	-2.4	-2.6
Numb of Excess	730	631	*552*	468	412	363	330
Robust ξ	0.201	0.225	*0.259*	0.220	0.242	0.266	0.352
Robust σ	1.311	1.323	*1.309*	1.432	1.439	1.446	1.331
Return level (p=0.02)	-9.191	-9.895	*-10.663*	-10.890	-11.580	-12.345	-13.803
Rank	**weight**	**weight**	**weight**	**weight**	**weight**	**weight**	**weight**
26	0.9769	1	*1*	1	1	1	1
25	0.9743	1	*1*	1	1	1	1
24	0.9632	1	*1*	1	1	1	1
23	0.9293	1	*1*	1	1	1	1
22	0.8928	1	*1*	1	1	1	1
21	0.8843	1	*1*	1	1	1	1
20	0.8479	1	*1*	1	1	1	1
19	0.8366	0.9977	*1*	1	1	1	1
18	0.8335	0.9941	*1*	1	1	1	1
17	0.8107	0.9670	*1*	1	1	1	1
16	0.7828	0.9337	*1*	1	1	1	1
15	0.7322	0.8734	*1*	1	1	1	1
14	0.6568	0.7832	*0.9082*	1	1	1	1
13	0.6041	0.7200	*0.8360*	0.9208	1	1	1
12	0.5886	0.7014	*0.8146*	0.8948	1	1	1
11	0.5319	0.6334	*0.7368*	0.8000	0.9208	1	1
10	0.5261	0.6264	*0.7288*	0.7903	0.9095	1	1
9	0.4886	0.5815	*0.6772*	0.7279	0.8373	0.9591	1
8	0.4458	0.5301	*0.6182*	0.6569	0.7550	0.8642	1
7	0.4453	0.5295	*0.6175*	0.6561	0.7541	0.8631	1
6	0.4339	0.5159	*0.6019*	0.6374	0.7324	0.8380	0.9899
5	0.3914	0.4651	*0.5435*	0.5681	0.6520	0.7451	0.8888
4	0.3707	0.4404	*0.5150*	0.5346	0.6132	0.7002	0.8396
3	0.2830	0.3359	*0.3948*	0.3961	0.4529	0.5153	0.6344
2	0.2615	0.3105	*0.3655*	0.3631	0.4149	0.4717	0.5852
1	0.2265	0.2690	*0.3176*	0.3100	0.3540	0.4019	0.5057

Table 5.13: Predictive discrepancy analysis

Samples		$D_{\hat{\theta}}$	D_{avg}	D_D	DIC
Simulated Data	Asymmetry	21746	21753	7	21760
	Symmetry	21778	21782	4	21786
IBM Return	Asymmetry	5448.2	5454.8	6.6	5461.4
	Symmetry	5454.2	5457.8	3.6	5461.4

model approached with the true parameter value $\theta = (0, 2, 2, 0.05, 1.3, -4, 0.4, 1.4)$ and size 3000. Then the IBM daily return (from 30/08/1993 to 22/05/2002) is used in testing the asymmetry distribution in tails. The results are reported in Table 5.13. The table reports the deviance results for both fitting by asymmetric and symmetric models for all samples. The D_D for asymmetric model fitting are larger than it from the symmetric model indicating the extra parameters added in asymmetric model are effective. The D_D can be generally be thought of as the number of "unconstrained" parameters in the model, where a parameter counts as: *1* if it is estimated with no constraints or prior information; *0* if it is fully constraint or the parameter fully depends on prior information; or an intermediate value if both the data and prior distributions are informative. Hence, the difference between D_D from the number of parameters in the model represents the constraint from the prior information associated. The simulated sample is asymmetric in tails and consequently has a lower DIC for asymmetric model fitting.

The D_D for IBM under asymmetric model is larger than it under symmetric model and a lower DIC for asymmetric model. The difference for DIC under the simple and complex model is not significant for the IBM data suggesting a similar performance in overall. However, in terms of the effect on tails, the difference can be significant. Figure 5.10 gives the posterior quantiles distribution under both asymmetry and symmetry model with the p-value (tail probability), which is defined as $min\{P(q(x^*) \leq q(x), P(q(x^*) \geq q(x)\}$. $q(x^*)$ represents the quantile from the samples from posterior predictive distribution and $q(x)$ is the quantile from the sample. Therefore, a large p-value indicates a good model fitted as the sample quantile lies close to the mode of the posterior predictive distributions of the test statistics. The p-value from asymmetric model for those four extreme quantiles *(0.001,0.01,0.99,0.999)* are all larger than the p-value from the symmetric model shows that the asymmetric model describes the tails better than the symmetric model. The results for the bulk quantiles are mixed, which induces the insignificant difference in overall model deviance analysis. The lack of significance in the bulk is unsurprising as the bulk distribution if symmetric, and fit in the bulk will not be substantially impacted by fit in the tails. The asymmetric mixture model is still preferred with the interests only lie in tails.

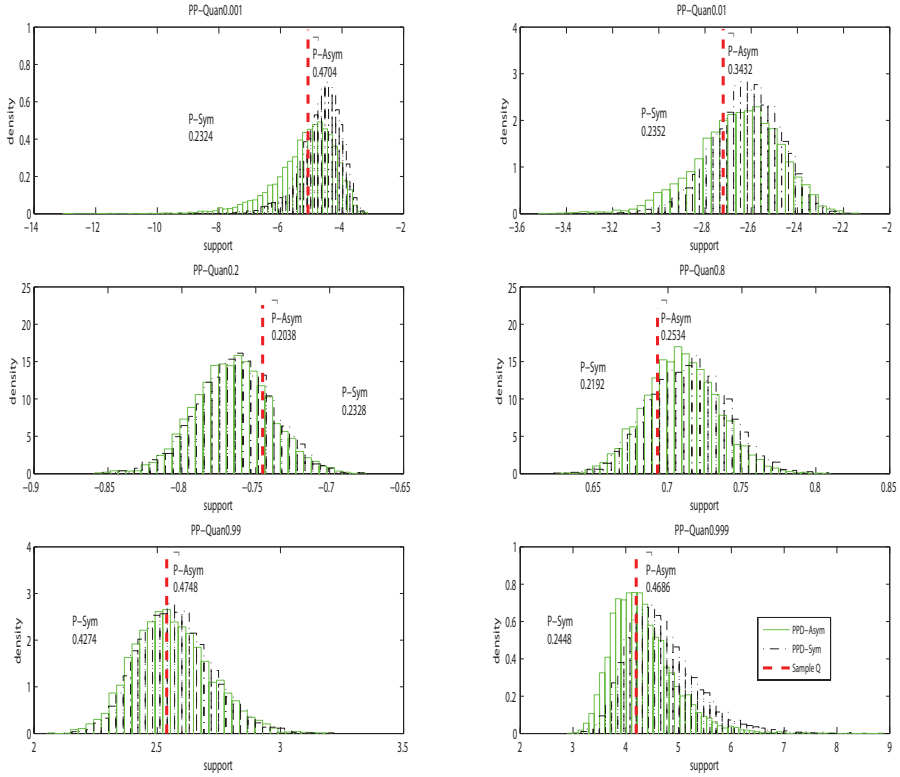

Figure 5.10: PPQ of Asymmetric and Symmetric models. P-Asym is the tail probability calculated based on asymmetric GNG and P-Sym is the tail probability calculated based on symmetric GNG. The histogram with solid line represents the PPQ distribution of asymmetric model and the histogram with dash-dot line indicates the PPQ distribution of symmetric model.

5.8 Summary

In this charpter, we propose a new GPD based mixture model, to overcome the challenge of threshold selection in traditional GPD applications. The proposed mixture model is able to account for the uncertainty associated with the threshold choice in inferences, as the threshold is an explicit parameter of the model to be estimated. As the threshold is estimated as part of the inference process, the model fitting (including threshold choice) is easily automated for large scale application to multiple time series which is very challenging for the traditional approaches to GPD threshold choice.

The model is able to extrapolate both the upper and lower tails of the distribution simultaneously, which can be more applicable in financial applications than mixture models with only a single GPD tail. The proposed mixture model is also very flexible in the tail features as they can capture all sorts of tails, e.g. either heavy or non-heavy tails (alternatively fat/thin tails), permitting both symmetric and asymmetric tail behaviours in the upper/lower tails.

A simulation study has shown the performance of a Bayesian inference approach for fitting the new GPD mixture model. The mixture model was also applied to various different population distributions (symmetric and asymmetric) and was shown to provide good approximations to various high quantiles, and in particular being only slightly less efficient compared to using the correct model for estimating the corresponding quantiles. The latter simulation study also demonstrates the general applicability of the proposed modelling approach.

The choice of the distribution used to capture the main mode of the distribution in the mixture model (normal considered in this paper) was shown in the simulation study to affect the performance of the model in capturing the high (or low) tail quantiles. The normal distribution is suggested for the financial applications in this thesis, due to their inherent unimodal, approximately symmetric and quadratic shape around the mode. However, for applications where an asymmetric mode is expected alternative distributions for the bulk should be considered, e.g. a Weibull or gamma distribution. Alternatively, a natural extension for the bulk distribution would be a mixture of uniform distributions extending the approach of Tancredi, Anderson, and O'Hagan (2006).

Chapter 6

Extreme Modelling in Value-at-Risk

6.1 Introduction

Extreme value models have been widely used to assess financial risk such as risk due to adverse market movements, see for example Embrechts, Klüppelberg, and Mikosch (1997). The current financial crisis, as with those of 1990's, has further stimulated interest in describing the probability of such extreme events Gençay, Selçuk, and Ulugülyağci (2003). Extreme value models describe the stochastic dynamics of a process for states with small chances of realization, and typically beyond the range of observed data Beirlant, Goegebeur, Segers, and Teugels (2004). They are therefore suitable for capturing tail related quantities for risk measurement and control. Value-at-Risk (VaR) is one such risk measure, which quantifies the largest possible loss-and-gain of a portfolio over a fixed holding period for a given low probability (see Duffie and Pan (1997) and Jorion (2000) for a variety of definitions and expansions of VaR). In statistical terms, the VaR is estimated as the extreme quantiles of returns due to unexpected market shortfalls. Three common statistical approaches to estimate VaR (McNeil and Frey 2000; Gençay, Selçuk, and Ulugülyağci 2003) are non-parametric historical simulation methods, parametric methods based on econometric models with volatility dynamics and and various extreme value theory (EVT) based methods, where commonly the generalized Pareto distribution (GPD) is assumed for the tail distribution.

Non-parametric methods for estimating VaR are challenging due to the inherent lack of sample information in the tails of the distribution (Embrechts, Klüppelberg, and Mikosch 1997). Parametric methods such as the autoregressive conditional heteroscedasticity (ARCH) and generalized ARCH (GARCH) model, and their many variants, typically make the assumption of conditional normality for the residuals which is typically unreal-

istic as it is commonly observed that financial series display heavier tails. Extreme value theory (EVT) based models are based upon an asymptotic approximation for the tail distributions, which are very flexible in terms of the allowable tail shape behaviour. The attraction of the EVT based methods is that they can provide mathematically and statistically justifiable parametric model for the tails of distribution which can give reliable extrapolations beyond the range of the observed data.

There are two issues in applying the traditional EVT based model, particular GPD, for estimating the VaR: the dependence of extremes (financial returns typically show clusters of observations in the tails) and the threshold choice of GPD (i.e. at which level of extremity into the tails of the data is the GPD a good model). The classical extreme value theory used to justify the GPD for capturing the tail of a distribution assumes the observations are independent and identically distributed.

There are several solutions to these issues, some of which are now discussed. Parametric financial models and the EVT theory have been combined in various forms to capture the impact of the heteroscedastic/dependence on the tail behavior of the return series. One approach is to introduce a time structure on the parameter of the GPD to capture the heteroscedastic process, see Bali and Weinbaum (2007) or Zhao, Scarrott, Oxley, and Reale (2010b) for example. However, among other difficulties with this approach is defining a stationary time series structure based on irregularly spaced time series. McNeil and Frey (2000) put forward a two stage approach by combining the GARCH model to estimate volatilities and then use the EVT based (GPD) to model the tail of the standard innovations of the GARCH above some chosen thresholds. After removal of the stochastic volatility, the standard innovations are approximately iid, permitting the use of the GPD as an asymptotically justified model for the tail.

A key problem with the above two stage method is the threshold selection for the GPD and accounting for the uncertainty due to the selection as discussed in Chapter 5. A mixture GPD model with threshold parameters can resolve this problem and supply more realistic uncertainty estimates. In many applications automated threshold estimation would be beneficial, e.g. in forecasting or applications with many financial series.

A mixture model, where both tails are GPD is developed in Chapter 5, which is suitable to capture both tails simultaneously and account for the uncertainty about the thresholds in any inferences made. Clearly this extension provides a very flexible model for capturing all forms of tail behaviour, potentially allowing for asymmetry in the distribution of the upper and lower tails. In this chapter, we are interested in quantifying the tail related VaR risk measure using the two stage McNeil and Frey (2000) approach, but with our two GPD tail mixture model defined in Chapter 5. More specifically, we define a mixture distribution with three components: a GPD for losses tail, a GPD for

gains tail and the normal for the innovations between main mode between the two GPD thresholds. The distribution selected for the non-extreme data can affect estimation of the tail distribution and therefore it is necessary to choose it according to the application. The normal distribution is suggested for the financial applications in this paper, due to their inherent unimodal, approximately symmetric and quadratic shape around the mode. A Bayesian method of estimation is used for the mixture model as it can take the advantage of any expert prior information, which is important in tail estimation due to the inherent sparsity of data. The proposed method is applied to the forecasting of VaR on daily return series of both individual stock and market index including the period of the most recent world financial crisis (year 2008).

Realized volatility has been recently widely documented as McAleer and Medeiros (2008), Oomen (2005) and Andersen, Bollerslev, Diebold, and Ebens (2001). Alternatively, we consider a VaR model with realized volatility in the first stage rather than GARCH volatility (latent volatility). As opposed to latent volatility, realized volatility is a model free estimate, using the cumulative variance of the intra-day returns which provides a good approximation of daily variations. In the purpose of forecasting, we assume a simple AR(1) in the log variance to predict the expected variance on the next day. The risk of unexpected falls (or adverse market movement) are still assessed by our two tail mixture model GNG. The VaR forecasted by using realized volatility will be shown to adjust faster compared to the VaR by latent volatility, as the realized volatility is more volatile than GARCH volatility. However the high frequency intra-day return data can be hard to acquire and filtering the large amount of transactions to get real transactions can be time consuming and affect the estimators of daily volatility.

The proposed two stage VaR estimator with our new mixture GPD model is described in Section 6.2. Section 6.3 describes the procedure for estimating VaR. Section 6.6 describes the data and gives the empirical results from application to returns from a stock and index. An alternative VaR with realized volatility is introduced in Section 6.6 with the empirical results. Conclusions and discussion are supplied in Section 6.7.

6.2 Two Stage VaR Model: GARCH-GNG

Let $\{R_t\}$ be a strictly stationary daily log return series on a financial asset at time t. The two stage approach to estimate the VaR is as follows:

1. Fit a GARCH form volatility model to $\{R_t\}$ and obtain the standardized innovation term x_t as $R_t = E(R_t) + v_t x_t$. Here, the $E(R_t)$ is the expected return at time t and v_t is the volatility estimator from a GARCH model. The form of GARCH can

be selected according to the particular application.

2. Fit the proposed GNG mixture model to $\{x_t\}$ (the standardized innovation sequence) as described in Chapter 5.2.2. The upper tail of the mixture model represent gains and the lower tail represents the losses.

The first stage GARCH model is fitted using a commonly used standard maximum likelihood method. It is worth noting here that estimation of the volatility model parameters does not suffer as in the extremal modelling, due to the lack of observations in the tails. However, the GNG mixture model is estimated using Bayesian inference, as the complexity of the likelihood for this model means it would be challenging to maximise directly and Bayesian inference also permits use of prior information which can substantially aid estimation of tail quantities (like VaR) due to the inherent paucity of sample information.

6.3 Estimate VaR

The GPD cumulative distribution function defined as Equation 5.5 is the probability for excess over a high threshold, a conditional probability of $P(X < x|X > u)$ for upper tail. The condition quantile, for excess, of the traditional fixed-threshold GPD can be estimated according to Equation 2.6. If we are interested in estimating the $1-p$ quantile (where p is small so an upper tail quantile) of the entire population distribution (otherwise known as the return level associated with a return period of $1/p$), e.g. in VaR, then we need to scale the conditional GPD distribution function by the probability of being above the threshold, as Equation 2.7 for the unconditional quantile of traditional GPD with a $p^* = P(X > u_r)$ giving:

$$p(X < x) = P(X < x|X > u_r)P(X > u_r) = G(x|\xi_r, \sigma_r, u_r)p^*.$$

In the traditional fixed threshold approach, $p^* = P(X > u_r)$ is estimated using the sample proportion above/below the threshold.

Since the proposed GNG mixture model describes the entire sample distribution, the quantile for both tails can be estimate more straightforward as it is the inverse of distribution function $GNG^{-1}(x)$. With Bayesian inference, the quantile can be obtained by posterior predictive distribution. Same as model parameter estimation, the mean of posterior quantile is the mean over the HPD interval of posterior predictive quantile distribution.

After obtaining the extreme quantile for unexpected fall or rise, the VaR with a return

period of $1/p$ at time t is the sum of the expected return and the shot rise or fall of returns given by:

$$E(R_{p,t}) = E(R_t) + v_t q_p(x). \tag{6.1}$$

When forecasting, the 1-step ahead prediction of the conditional quantile q_p for the upper tail is defined as:

$$R_{p,t(1)} = \inf\{F(R_t) \geq 1 - p|\varphi_{t-1}\} \tag{6.2}$$

where φ_{t-1} is the information up to day $t-1$ and $F(R)$ is the distribution function of return. A similar result is obtainable for the lower tail 1-step ahead forecasts.

6.4 Application to Risk Estimation during Financial Crises

We use the proposed new VaR method to quantify the risk of extreme returns during financial crisis periods and compare the method with McNeil and Frey's approach, using a fixed-threshold GPD for second stage, to show the benefits with the automated threshold estimation.

6.4.1 Data

We apply the fixed GPD and the proposed GNG based methodologies to the daily returns (log return) of the Citigroup and S&P100 index over the period 17 January 2001 to 30 December 2008. These include 2000 daily prices and two major financial crises, namely 9/11 and the developing 2008-2009 financial crisis. We use a GARCH(1,1) for the volatility model. After removing the conditional variance (volatility) the standardized innovations x_t are essentially an independent sequence such as showed by Figure 6.1 of S&P100, which is beneficial for inference purposes for the second stage which assumes independence. Results of Citigroup are similar and not shown for brevity.

It is clear from Figure 6.2 that both the Gaussian and student-t distributions do not adequately represent the tail innovation distributions of the Citigroup, as they are sufficiently inflexible at representing the tail shape. Both tails are heavier than the normal tail and lighter than the t distribution. The Q-Q plots for the S&P100 index gives similar results as the Citigroup, so are not shown for brevity. A model which provides more flexibility in the tails of the distribution of the standardized innovations is therefore

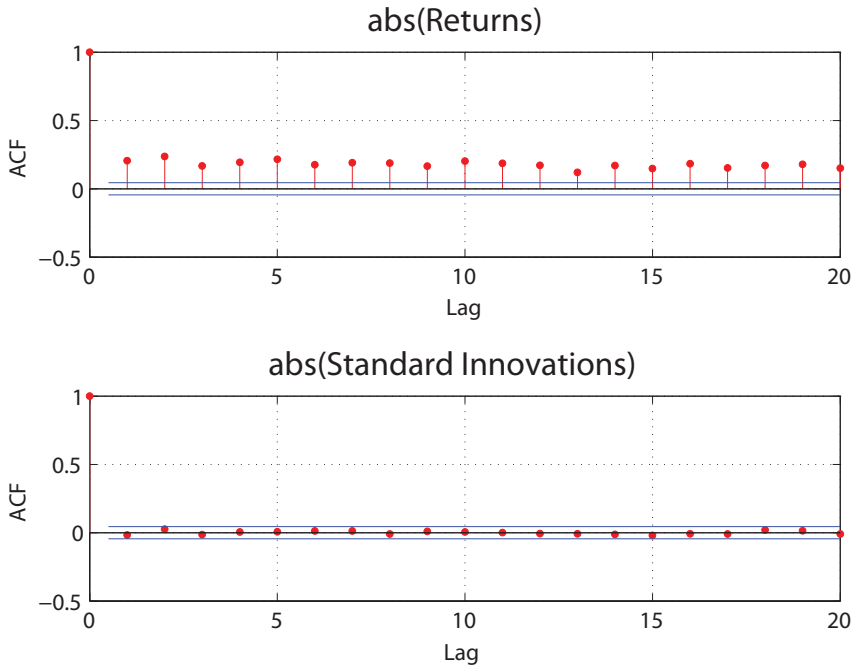

Figure 6.1: Autocorrelation function plot of S&P100: The first plot is ACF on $|R_t|$ and the second plot is ACF on $|x_t|$, the standardized innovation term.

required in VaR, particularly as it is the tails that are key for reliably estimating the expected fall or rise from the expected return.

6.4.2 Empirical Results

Estimation Results

The mixture GPD model (GNG) proposed in Chapter 5, along with the traditional fixed GPD approach, provide flexibility in capturing different tail behaviours. The model is suitable for any combinations of three types of tail behaviours with the typed determined by the shape parameter of the GPD, namely exponential type ($\xi = 0$), heavy tailed ($\xi > 0$) and short tailed ($\xi < 0$). The models also permits asymmetric tail behaviours (e.g. lower tail could be of exponential form and upper tail a heavy tail). The tails can be constrained to be symmetric if required, and various tests for evidence of asymmetry can be derived, see example in Section 5.7.

The proposed VaR model in Section 6.2 is applied to the return sequences of S&P100 and Citigroup. Figure 6.3 shows the fitted density of GNG model on the standardized residual sequence of Citigroup. The density of the mixture distribution is close to continues at the thresholds and shows asymmetry feature in tail distributions. The lower tail, representing the unexpected falls of return, has a relative heavier tail compare to

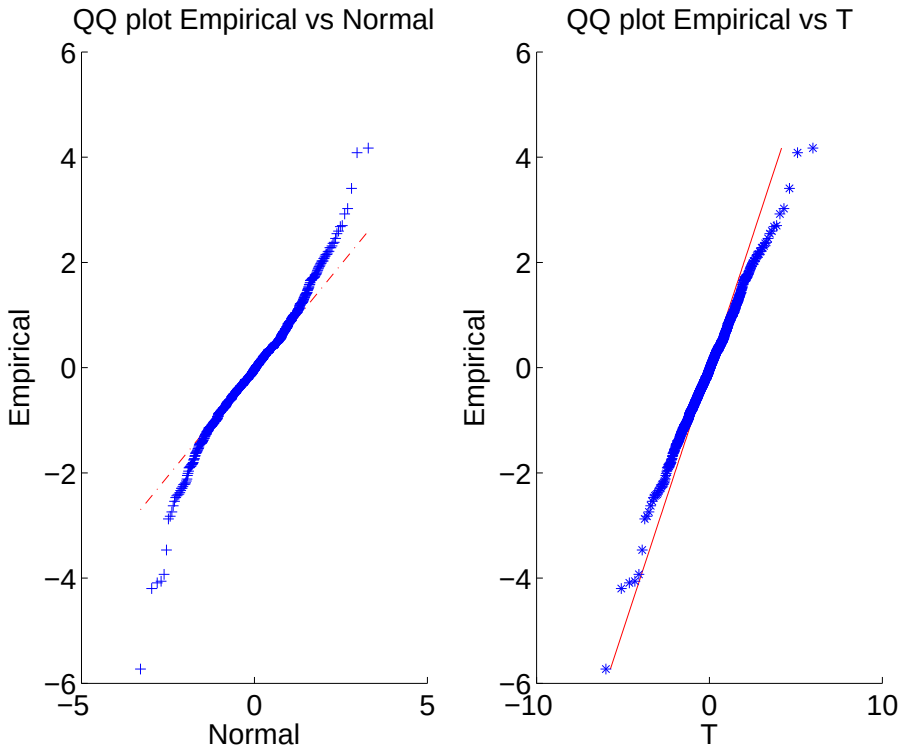

Figure 6.2: Quantile to Quantile plot of Citigroup. The left plot in the figure is the empirical quantiles versus the normal quantiles and the right plot is the empirical quantiles versus the t quantiles.

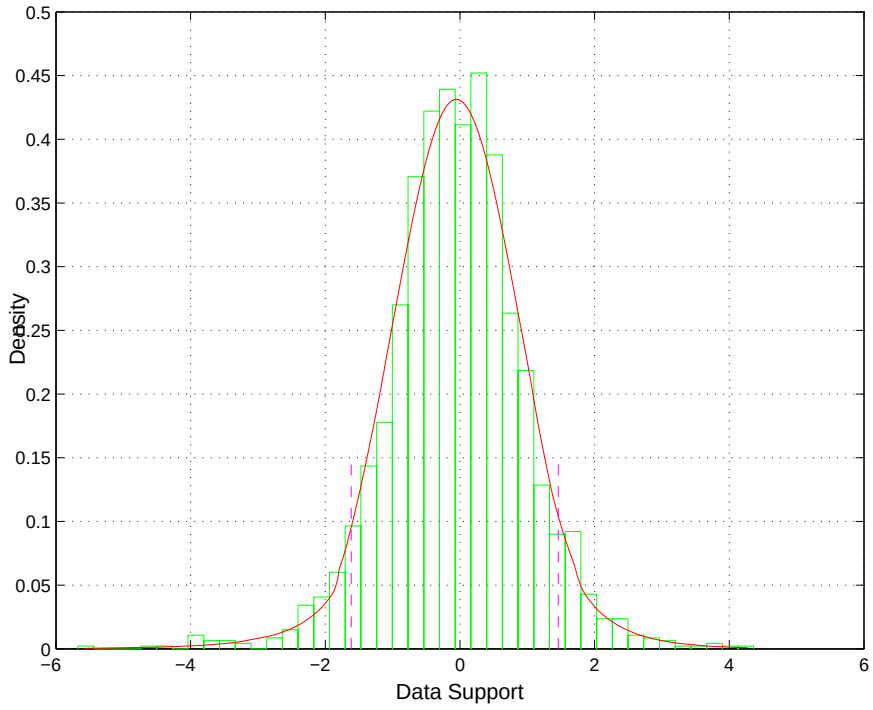

Figure 6.3: Posterior predicted density of GNG on the standard innovations of Citigroup. The two vertical dashed lines are the upper and lower threshold estimated by the model.

the upper tail.

Asymmetry vs. Symmetry

The return level (VaR quantile) plot is shown in Figure 6.4. Compare to the normal distributions in the figure, which is clear thinner in both tails than the sample, the return level based on GNG model indicates a precise extreme quantile estimate and the consistency with the empirical tail distributions. The credible interval of posterior predicted return level contains almost all sample quatiles.

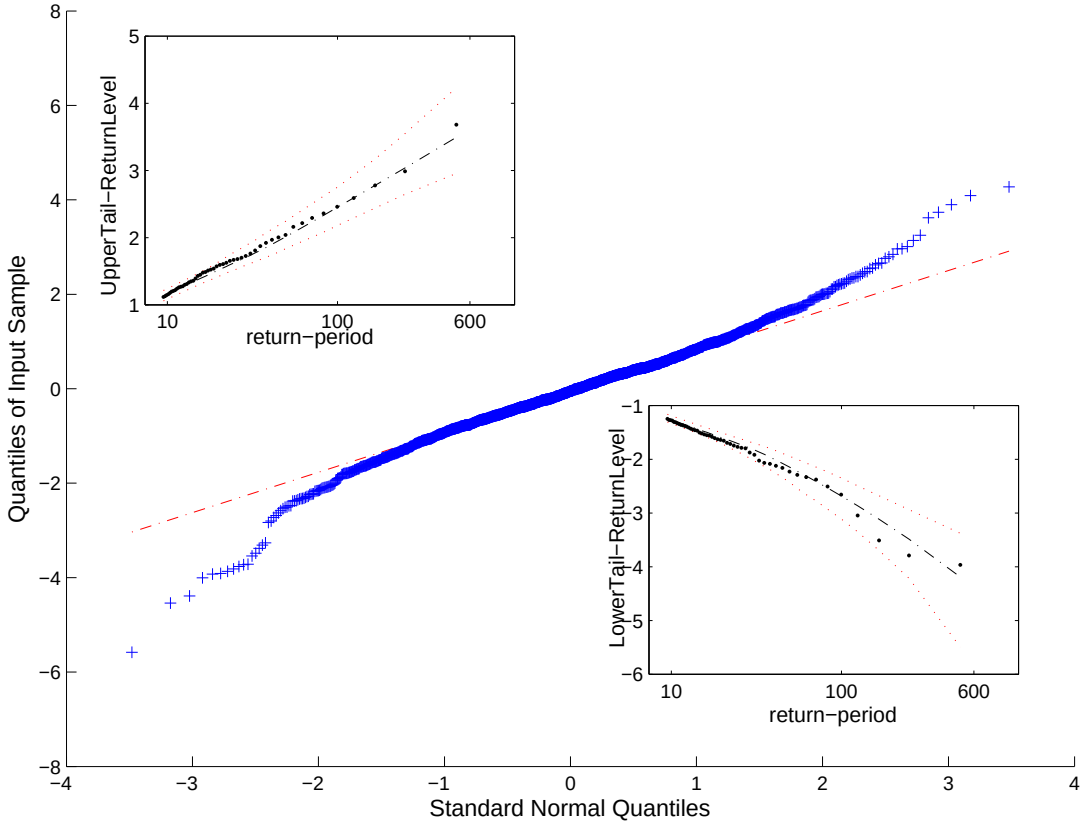

Figure 6.4: The estimated return level (VaR quantiles times 100) of Citigroup from asymmetric GNG model. Sample values are presented by dots, and posterior predictive estimates are shown by dashed lines with credible interval in doted line.

Further, we apply a symmetric two tail mixture model to Citigroup, assuming the upper and lower tails following a same GPD, and re-plot the return level to compare the estimates with asymmetric two tail mixture model. The return level plot of symmetric model is shown in Figure 6.5. The discrepancy between model values and actual observations is larger than that of asymmetric model, and not all observations fall into the credible interval. The results suggest that asymmetric mixture model a better approximation of higher quantiles of extremes and a more reliable uncertainty estimate.

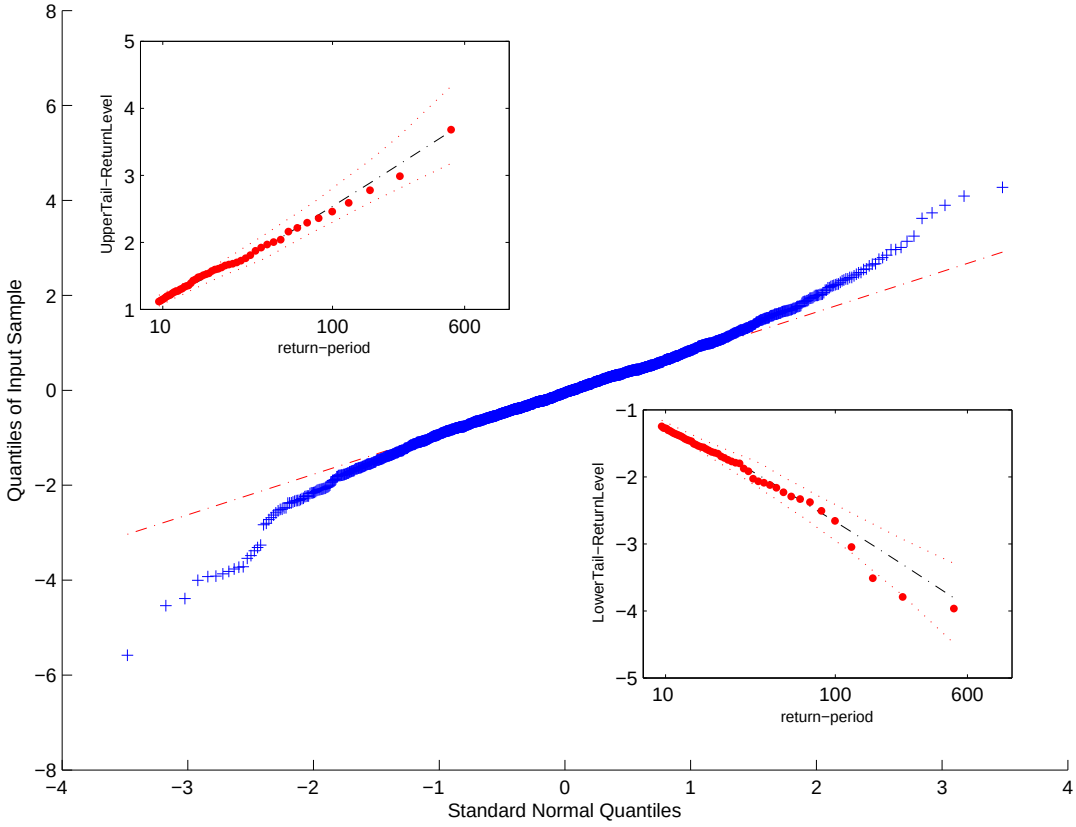

Figure 6.5: The estimated return level (VaR quantiles times 100) of Citigroup from symmetric GNG model. Sample values are presented by dots, and posterior predictive estimates are shown by dashed lines with credible interval in doted line.

Uncertainty of Threshold Choice

In order to demonstrate the problem of the threshold selection, we compare the estimation results of the two approaches in fitting the innovations of these financial time series. We use the maximum likelihood estimation (MLE) for the fixed threshold GPD, since this is the common approach in the literature and there is no substantive difference between results from the Bayesian method and MLE for fixed GPD when there is no strong expert opinion of the extreme distribution. The use of relatively diffuse priors in the Bayesian inference for the GNG model means the results are directly comparable.

Table 6.1 illustrates the effect of threshold selection in the traditional fixed threshold GPD approach. The GPD is fitted to the standardized innovation term of the Citigroup data with different thresholds and compared to the GNG mixture model. The choice of the threshold value results in variation of the GPD shape and scale estimators, and consequently for estimates of the VaR. Even though these variations appear rather slight for this data, the differences can have a substantial impact on the quantile and VaR estimates.

Table 6.1: Parameter estimates for both the fixed threshold GPD conditional on different threshold values and the GNG (two tail mixture GPD) models. In the table, POE is the proportion of innovations above/below the upper/lower thresholds respectively. u denotes the threshold value, ξ is the shape and σ is the scale of GPD.

Upper Tail	*POE*	u	ξ	*CI*		σ	*CI*	
	0.2	0.6716	-0.0607	-0.1498	0.0284	0.6691	0.5862	0.7638
Conditional GPD	0.15	0.8827	-0.0194	-0.1336	0.0949	0.6076	0.5173	0.7135
on fixed threshold	0.1	1.1178	-0.0429	-0.1769	0.0912	0.6296	0.5192	0.7634
	0.05	1.5721	0.079	-0.1721	0.33	0.5054	0.3677	0.6947
GNG model		1.4604	0.0304	-0.1897	0.2633	0.5991	0.4163	0.8012
	*C*I of u	0.9612	1.9182					
Lower Tail	*POE*	u	ξ	*CI*		σ	*CI*	
	0.20	-0.7788	0.0438	-0.0566	0.1443	0.6255	0.5437	0.7197
Conditional GPD	0.15	-0.9649	0.0505	-0.0661	0.1671	0.6246	0.5309	0.7347
on fixed threshold	0.10	-1.2481	0.1177	-0.0435	0.2790	0.5655	0.4575	0.6991
	0.05	-1.6410	0.1165	-0.1376	0.3705	0.6240	0.4531	0.8594
GNG model		-1.6149	0.1644	-0.1172	0.4323	0.6146	0.3909	0.9337
	*C*I of u	-1.9032	-1.1093					

Although, the Bayesian credible interval *(CI)* and MLE based confidence intervals have philosophically different interpretations, they are to some degree heuristically comparable due to the diffuse priors using in the Bayesian inference. The credible intervals for the shape and scale parameters of the GNG model, are wider than the confidence interval of the fixed threshold GPD as expected, as this in part comes from the extra source of threshold uncertainty. For heavy tailed distributions, a wider confidence inter-

val is expected since the uncertainty about the threshold is much higher compared to the short tailed distributions, where there is a natural end-point for the threshold value.

Table 6.2: Expected quantile estimators for the fixed-threshold-GPD and the GNG (two tail mixture GPD) models, along with credible intervals for the GNG model.

Upper Tail	**Upper Quantiles**				
Fixed GPD	u	**0.9**	**0.99**	**0.999**	**0.9999**
	0.6716	1.1257	2.5043	3.7029	4.7451
	0.8827	1.1281	2.4857	3.7841	5.0259
	1.1178	1.1178	2.4983	3.749	4.8823
	1.5721	1.2312	2.4395	3.8887	5.6269
GNG	u	**0.9**	**0.99**	**0.999**	**0.9999**
	1.4604	1.1359	2.4460	3.9691	5.6174
	CI	*(1.061,1.182)*	*(2.176,3.162)*	*(3.283,6.784)*	*(3.886,13.570)*
Lower Tail	**Lower Quantile**				
Fixed GPD	u	**0.1**	**0.01**	**0.001**	**0.0001**
	-0.7788	-1.2190	-2.7812	-4.5093	-6.4209
	-0.9649	-1.2208	-2.7774	-4.526	-6.4903
	-1.2481	-1.2481	-2.7438	-4.7053	-7.2776
	-1.641	-1.2254	-2.7456	-4.7333	-7.3323
GNG	u	**0.1**	**0.01**	**0.001**	**0.0001**
	-1.6149	-1.2376	-2.6779	-5.0217	-8.8012
	CI	*(-1.307,-1.165)*	*(-3.097,-2.343)*	*(-7.405,-3.829)*	*(-21.968,-4.720)*

Table 6.2 reports the estimated expected quantiles of the standardized innovations for the threshold choices in Table 6.1, to show the impact of threshold choice on the quantile estimators. The expected quantiles of the fixed threshold GPD are calculated as the return level of the GPD explained in Section 6.3. The expected quantiles for the GNG are calculated as the mean of the posterior predictive quantiles. It is clear that the expected quantile values vary when the threshold changes and the differences among them gets larger for quantiles further out into the (upper and lower) tails of the distribution. As mentioned previously, these differences tend to be magnified for heavier tailed distributions, which are commonly observed in financial data. Different choices of threshold can clearly lead to very different estimates of the extreme quantiles used in risk measures like VaR. The GNG explicitly estimates both thresholds at the same time, and accounts for the uncertainty in the estimates.

The variation considered by the GNG model represents the full uncertainty for all the model parameters. Figure 6.6 gives the sample quantiles from the predictive posterior quantile distributions based on the GNG estimates for the Citigroup data. As expected, the posterior quantile are more skewed for the higher quantiles, due to the reduced information in the tails. The predictive posterior for the upper tail are right skewed summarizing the lack of information for higher levels, and vice-versa for the lower tail

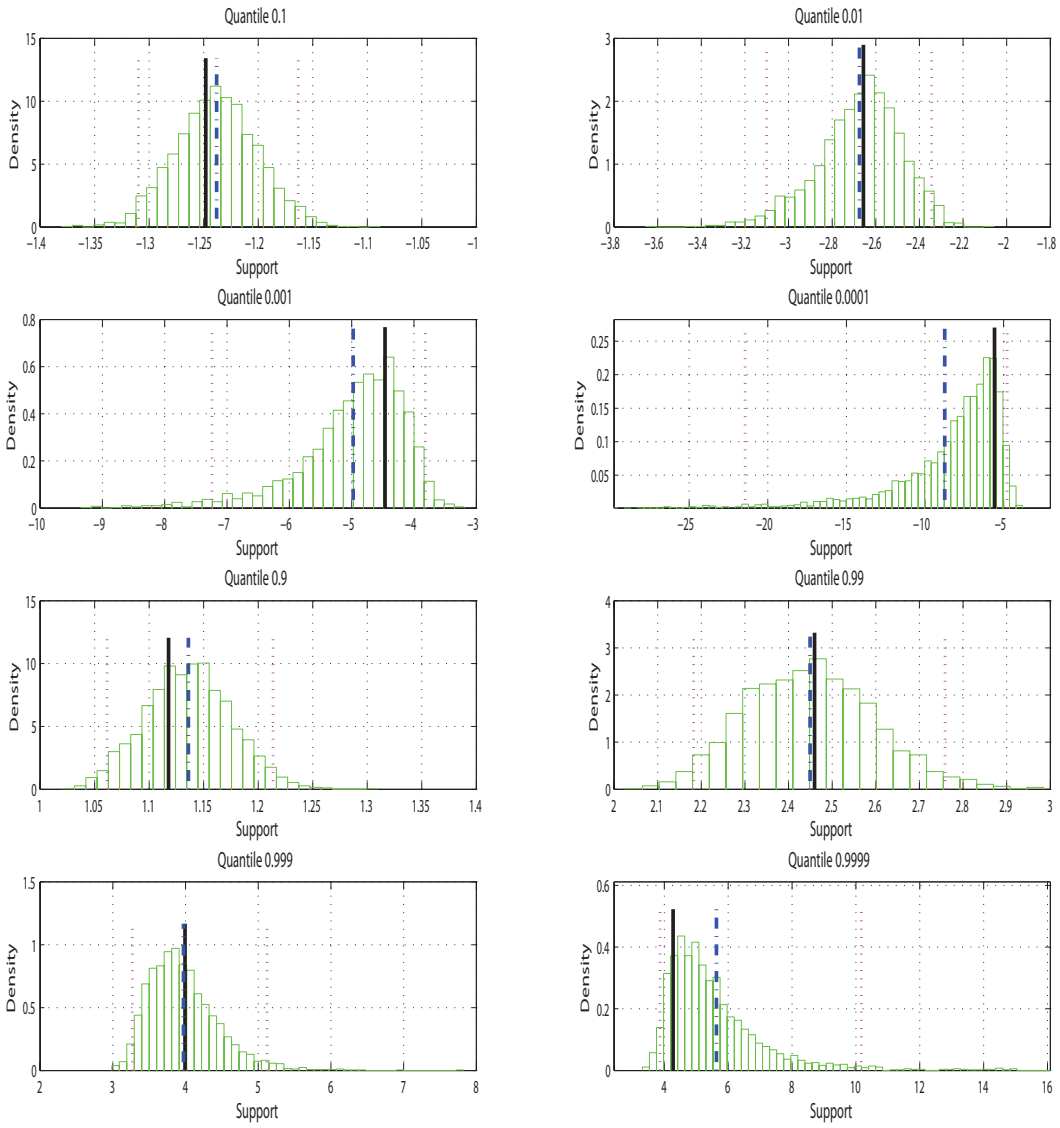

Figure 6.6: Posterior predictive quantile distributions for the Citigroup 2001-2008 innovations. The dashed line is mean of posterior quantiles and the sample quantile denoted by solid line. The interval between dotted line is the credible interval.

quantiles. Notice that the sample quantiles are included for comparison and are positioned at the mode and well within the credible interval, which demonstrates that the predictive quantile distributions are consistent with the sample information.

6.4.3 Forecasting VaR

Procedure

We use the proposed two stage method to produce the 1-step forecasts of daily return quantiles R_t^q for S&P100 index and Citigroup. Suppose the return sequence is $\{R_1, ..., R_t, ...R_T\}$, where $t < T$. In providing the forecast we only use historical information from n=1000 daily returns, i.e. approximately 4 years. The forecasted return quantile $\hat{R_t^q}$ is then dependent on all the information up to $t-1$.

The scheme is as follows at time t

1. Apply GARCH(1,1) model on $R_{t-n-1}, ..., R_{t-1}$ to obtain the standardized innovation sequence $x_{t-n-1}, ..., x_{t-1}$.

2. Obtain the one-step forecasting of the expected return $E(\hat{R_t})$ and the volatility $\hat{v_t}$ based on the estimators of step 1.

3. Apply the GNG model on the standardized innovation sequence $x_{t-n-1}, ..., x_{t-1}$, and forecast the expected quantile $\hat{x_t^q}$ based on the predictive posterior quantile distributions.

4. The forecasted return quantile can be calculated as $\hat{R_t^q} = E(\hat{R_t}) + \hat{v_t}\hat{x_t^q}$, using the obvious notation developments from Equation 6.1.

Repeat all the steps for each time point to obtain a sequence of conditional forecasted return quantiles $\{\hat{R_t^q}\}$ for the forecasting period. These forecasted return quantiles are termed "conditional" as they are conditional on the variance being assumed known, using the estimates from the GARCH. We also estimate the unconditional quantiles for comparison, which are the quantile estimators by applying the GNG model directly on the return series. The unconditional quantiles do not account for the dependence due to the volatility.

Results

Figure 6.7 gives the plot of real return observations of the Citigroup and forecasted conditional & unconditional quantiles for the 1% upper and lower quantiles. We have focused

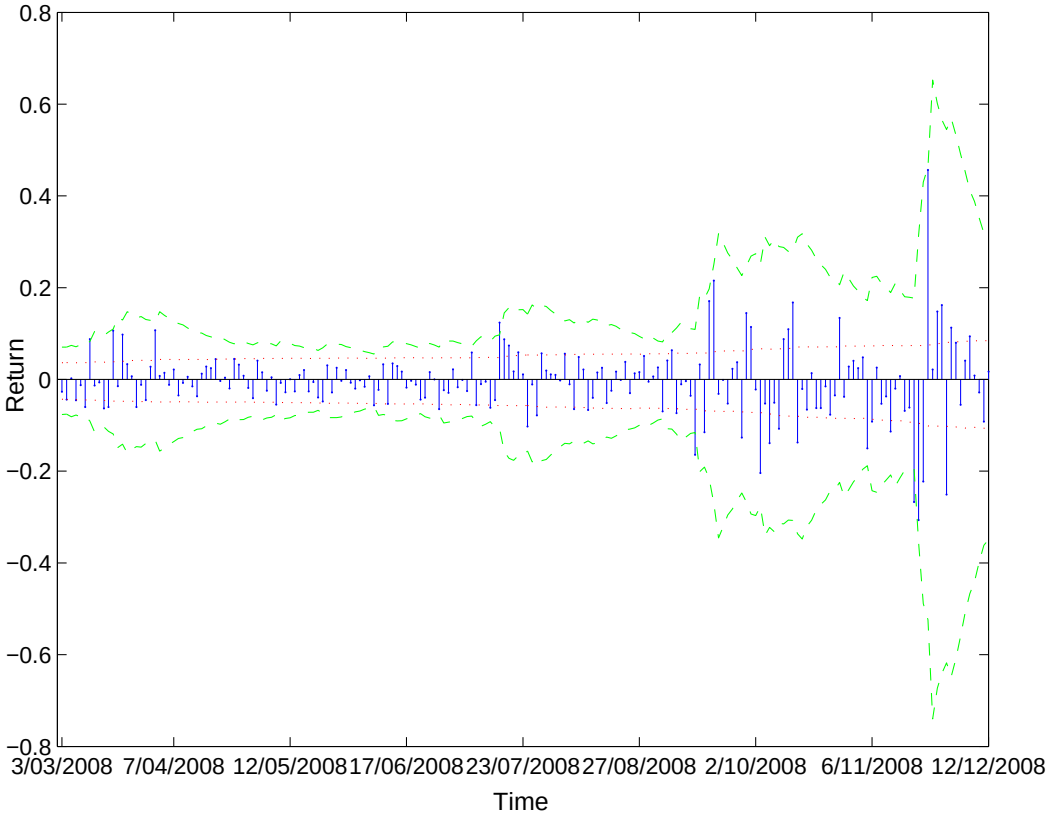

Figure 6.7: Conditional and unconditional return quantile forecasting of Citigroup 2008. The dashed line is the conditional quantiles and the dotted line denotes the unconditional quantile forecasted.

on the period from March 2008 to December 2008, as this covers the recent financial crisis. As expected, the unconditional quantile estimators fail to capture the changing volatility and dependence of the returns. The conditional forecasts, on the other hand, can describe both the dynamic of the conditional variance and the tails behavior of the expected falls and rises. The conditional approach based on the GARCH(1,1) is therefore clearly warranted due to it providing more realistic description of the return quantiles. Similar results were obtained for the S&P100 index with the focuss of the crisis of *9*.11 in 2001 as shown in Figure 6.8. These two example shows the importance of dependence structure of return in VaR and the predicted quantiles of VaR cannot adjust as fast as the market risk increase or decrease without the concern of volatilities.

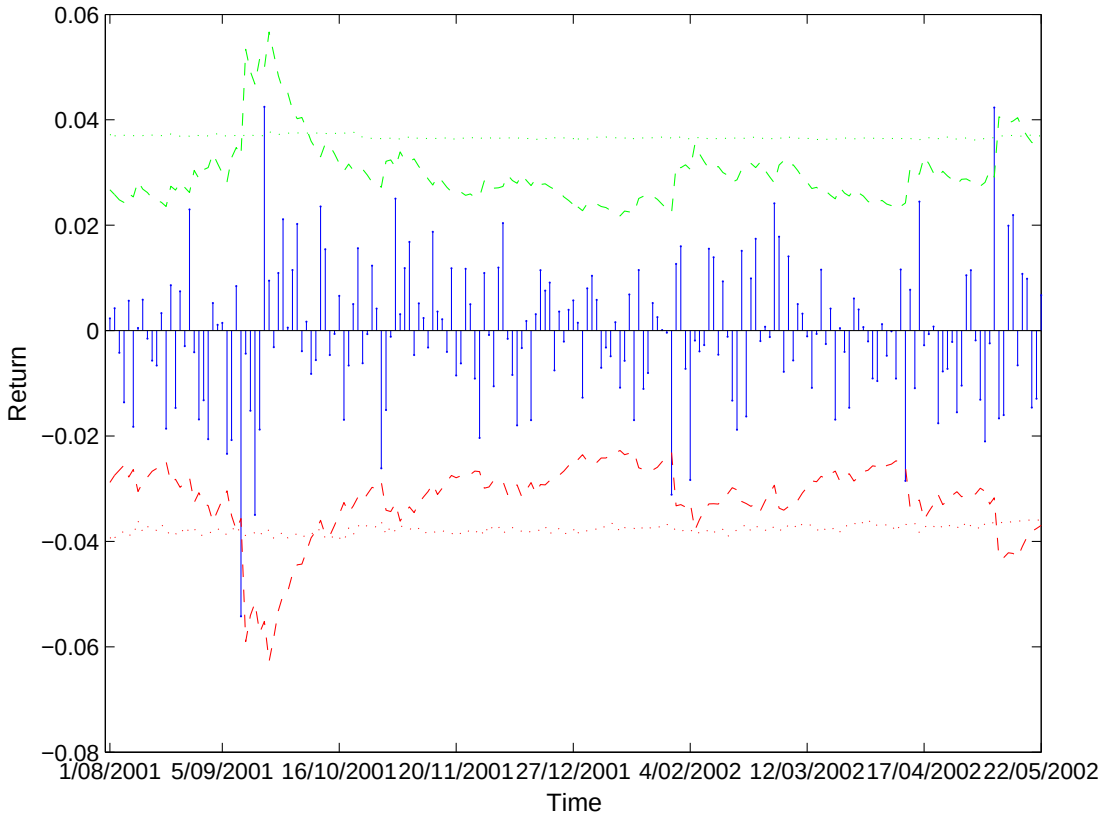

Figure 6.8: Conditional and unconditional return quantile forecasting of S&P100 2001. The dashed line is the conditional quantiles and the dotted line denotes the unconditional quantile forecasted.

A comparison between our method and the fixed threshold approach is now conducted. To implement this method in a forecasting framework it is necessary to choose the thresholds at each timepoint, which is clearly impractical. It is common in the literature to determine a threshold in advance, by specifying what proportion of the observations at each timepoint should above/below the threshold, using graphical diagnostics the details of which are not shown here for brevity. Figure 6.9 to Figure 6.12 show the comparison

between the GNG and the fixed threshold GPD for the upper and lower quantiles at the 0.5% level with associated confidence intervals. The proportion above/below the threshold for the fixed each threshold approach was fixed at 10%. Extensive comparisons with other sensible proportions above/below the thresholds provided no change in the conclusions drawn. Notice that the extreme quantiles for both tails based on the GNG are slightly larger than the fixed GPD based method, with a wider confidence interval as we expect due to accounting for the uncertainty about the threshold.

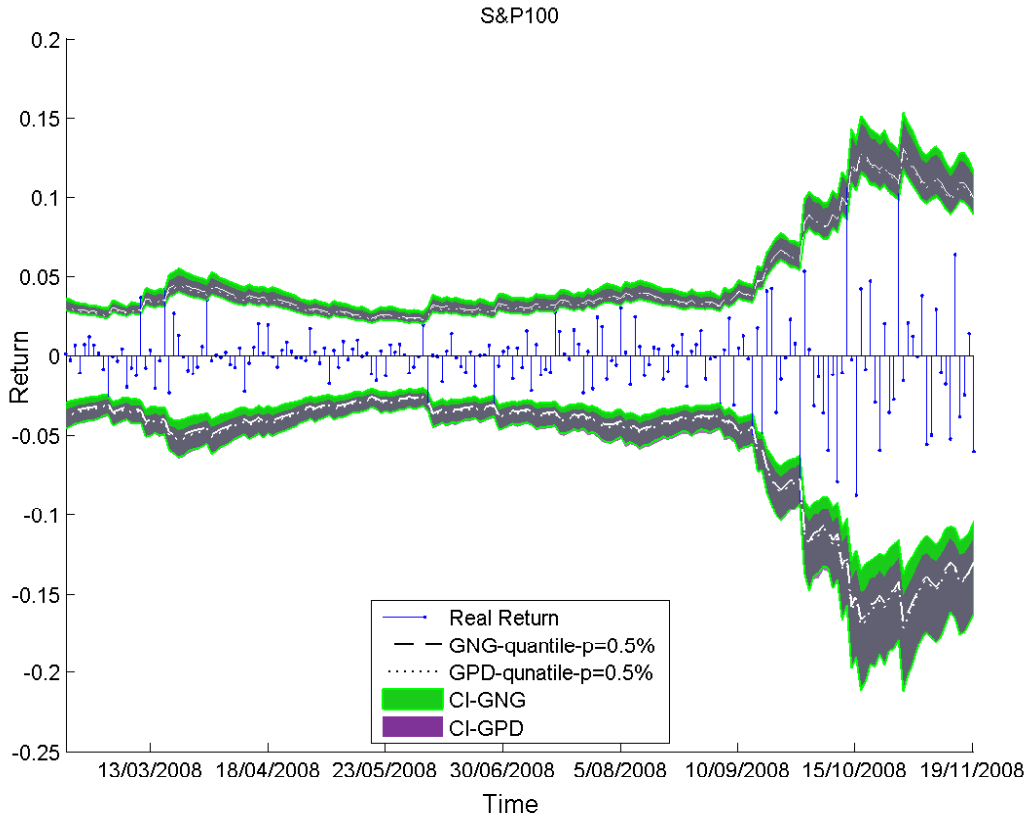

Figure 6.9: Conditional return quantile forecasting of S&P100 2008. The dashed line is the conditional quantiles based on GNG model and the interval in green is the uncertainty interval associated. The dotted line denotes the conditional qunatile forecasted based on the classical GPD model and the interval in purple represents the uncertainty interval (CI) with it.

The similarity of the estimates and the only slightly larger credible intervals is extremely pleasing, as the proposed methodology has not required a-priori specification of the threshold which is a major advantage over the traditional fixed GPD based method.

An interesting feature of the credible intervals for the mixture model approach is that they tend to be somewhat wider for heavier tails, such as shown in Figure 6.9 and Figure 6.10 by the larger credible intervals for the losses compared to the gains. The shape parameter estimated for the losses (not shown for brevity) is generally larger than

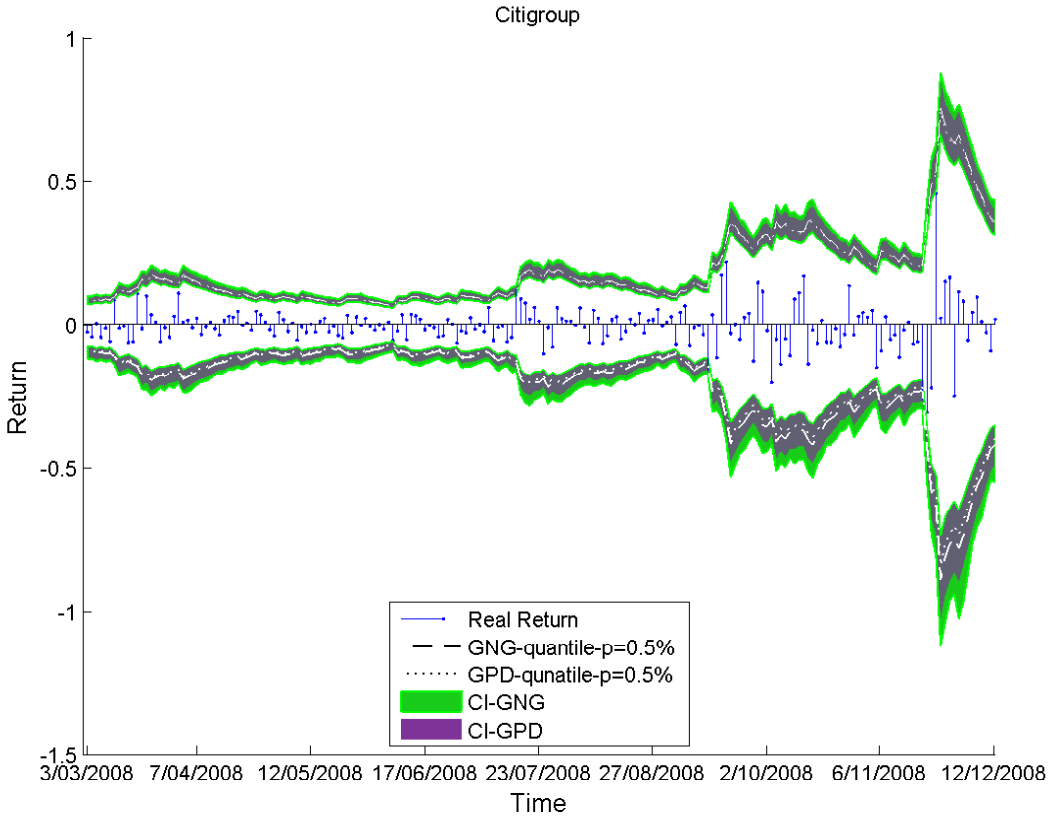

Figure 6.10: Conditional return quantile forecasting of Citigroup 2008. The dashed line is the conditional quantiles based on GNG model and the interval in green is the uncertainty interval associated. The dotted line denotes the conditional qunatile forecasted based on the classical GPD model and the interval in purple represents the uncertainty interval (CI) with it.

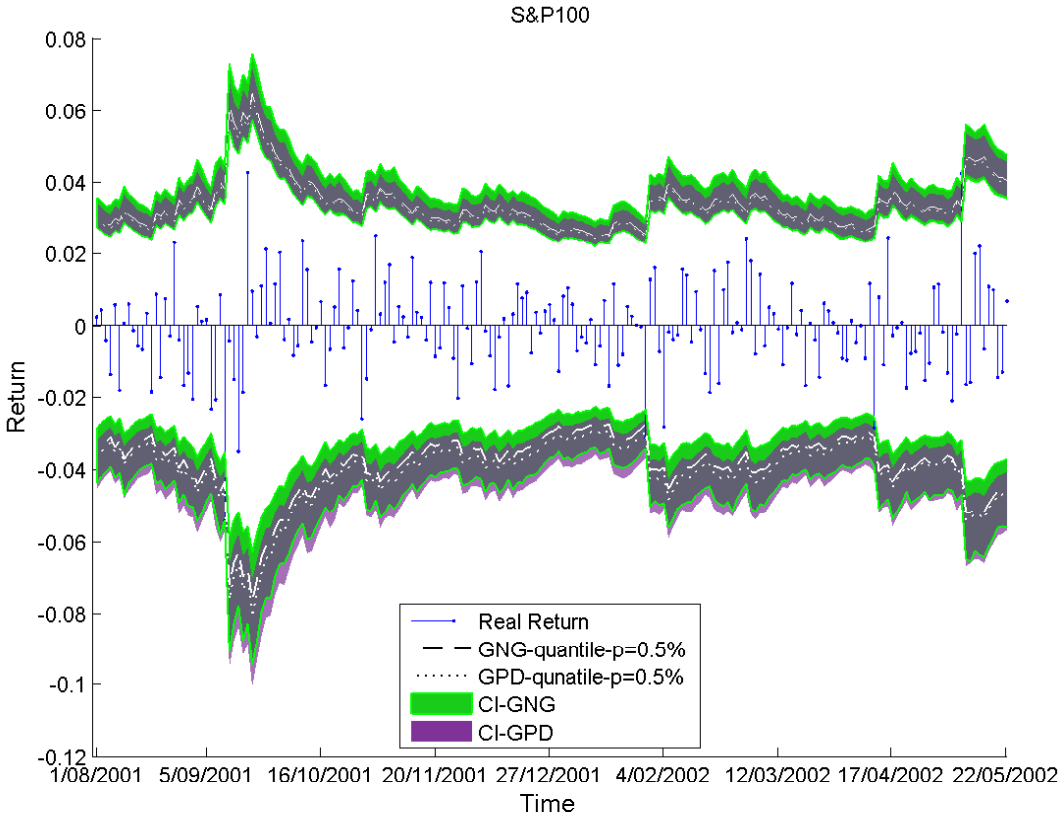

Figure 6.11: Conditional return quantile forecasting of S&P 2001. The dashed line is the conditional quantiles based on GNG model and the interval in green is the uncertainty interval associated. The dotted line denotes the conditional qunatile forecasted based on the classical GPD model and the interval in purple represents the uncertainty interval (CI) with it.

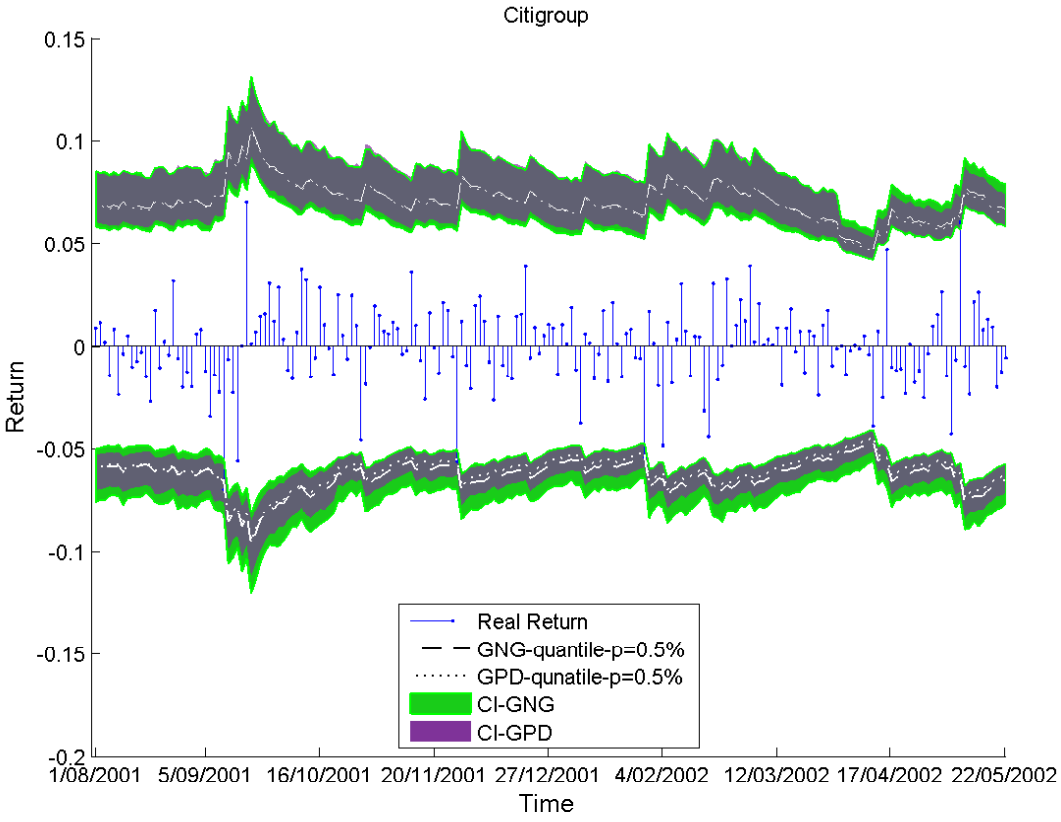

Figure 6.12: Conditional return quantile forecasting of Citigroup 2001. The dashed line is the conditional quantiles based on GNG model and the interval in green is the uncertainty interval associated. The dotted line denotes the conditional qunatile forecasted based on the classical GPD model and the interval in purple represents the uncertainty interval (CI) with it.

for the gains. This result implies that the uncertainty of threshold selection is likely to be more important for the heavy tail distributions relative to the light or short tails, which is commonly the case for financial data (particular for financial crises).

Notice from these figures that the larger credible intervals for the GNG model based quantile estimates, have better coverage of the actual returns than for the fixed threshold approach. The lower bound of the lower tail credible interval is approximately the same for both models. However, the upper bound of the lower tail credible interval is closer to the mode. For heavier tails (as in the losses in this example), the threshold related uncertainty appears as more uncertainty nearer the mode of the distribution, rather than further out into the lower tail. In contrast the gains credible interval for the GNG model extends further out into the upper tail than the corresponding interval for the fixed threshold model, whereas the lower bounds are very similar. The gains have a short tail (generally a negative shape parameter), hence the uncertainty associated with estimating the threshold leads to more uncertainty about the upper tail.

The alternative estimation by Bayesian inference for the fixed threshold GPD also has been used to see the difference in comparison the forecasted extreme quantiles with the estimation by MLE. Generally the results are very similar as previous results of MLE on the classical GPD, for example Figure 6.13 is very similar as Figure 6.9. We do not show all the results for brevity.

To summarize the above empirical results the proposed GARCH-GNG model based method of forecasting has the following advantages over the traditional fixed threshold GPD approach:

1. The model resolves the difficulty in threshold selection by treating the threshold as a further parameter to be estimated. This advantage becomes extremely beneficial for problems where automated application is required, as in the above forecasting example and applications containing a large number of series to apply the approach to. The automated threshold estimation also removes the subjectivity employed in the common methods used for threshold choice (e.g. using graphical model fit diagnostics like the mean residual life plot).

2. The model explicitly accounts for the uncertainty due to the threshold choice in VaR. Sequentially, the model supplies a more objective approach in predicting the extreme short fall and rise in the VaR along with a more precise uncertainty interval associated with the unexpected returns.

3. The model fits to both the gain and loss tails simultaneously which is more natural and convenient when the interest lies in both tails. The uncertainty estimates

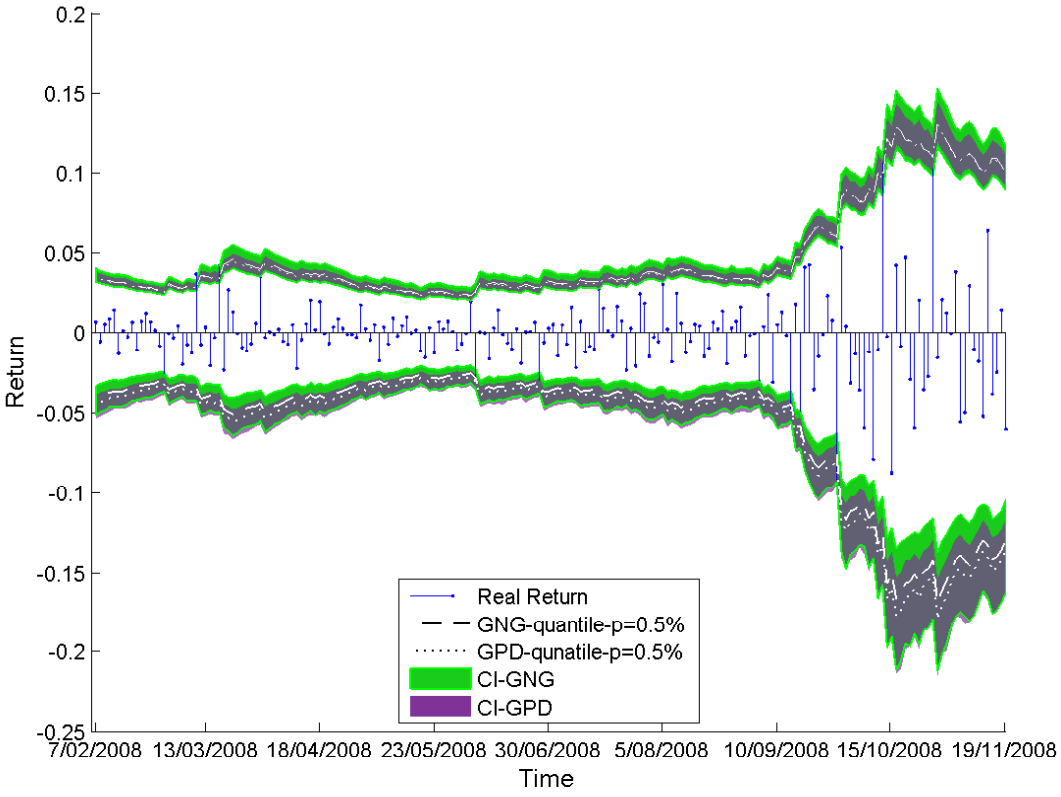

Figure 6.13: Conditional return quantile forecasting of S&P100 2008 (both model under Bayesian estimation). The dashed line is the conditional quantiles based on GNG model and the interval in green is the uncertainty interval associated. The dotted line denotes the conditional qunatile forecasted based on the classical GPD model and the interval in purple represents the uncertainty interval (CI) with it.

for the gains and losses are also more appropriate than if they were estimated separately for each tail, as is common practice.

6.5 Comparison on different methods in forecasting VaR

Section 6.4.3 shows some empirical results of using our developed GARCH-GNG model to forecast VaR, in particular during the financial crisis period. In this section we compare different approaches in VaR by examining the number of violations, which occurs whenever the forecasted extreme quantile is not able to cover the actual return level on the day $t+1$, such as $\hat{R}^q_{t+1} > R_{t+1}$ for negative returns or $\hat{R}^q_{t+1} < R_{t+1}$ for positive returns. A good model is the one can predict the extreme quantiles $\hat{R}^q_{t+1}$ with a risk of fail rate equals $p = 1 - q$ accurately. In terms of the number of violations, this means the observed number of violations is close to the expected number of violations based on p value. The methods used in comparison include GARCH-GNG model, GARCH-GPD model, GARCH-normal and GARCH-t. We apply those methods on four historical series of log returns for *4000* trading days: the Standard and Poors index (S&P100) from January 1993 to December 2008, the Citigroup (CITI) from January 1993 to December 2008, the Phoenix Technologies Ltd. (PETC) from May 1994 to March 2010, and the JPMorgan Chase & Co. (JPM) from May 1994 to March 2010.

The forecasted VaR $\hat{R}^q_{t+1}$ is conditional on the historical series $R_{t+1-m}, ..., R_t$. For all four methods, m is set to be *1000* which represents a period less than four years of data. Therefore the initial period of data, *1000* observations, is used to predict the VaR on the first following day and a moving window of *1000* observations is used from then on. This results in a forecasting period of *3000* days. In a such long period, it is not feasible to examine the fitted model day by day for each of these VaR methods. Particularly, GARCH-GPD is conditional on the threshold selection. It is not feasible to check it by graphical techniques and model diagnostics day by day. Following McNeil and Frey (2000) and after checking the model fitting for a few number of days, we fix the threshold at 90% empirical quantile for upper tail and 10% empirical quantile for lower tail. On each day t, we fit a new AR-GARCH model, with both normal and t innovations, and then fit the proposed GNG model and standard fixed threshold GPD. The forecasted extreme quantiles $\hat{R}^q_{t+1}$ can be calculated based on the different models. The number of violations is then counted for these quantiles for $q \in (0.005, 0.01, 0.03, 0.05, 0.95, 0.97, 0.99, 0.995)$.

Too many violations corresponds to underestimation of the VaR and too few viola-

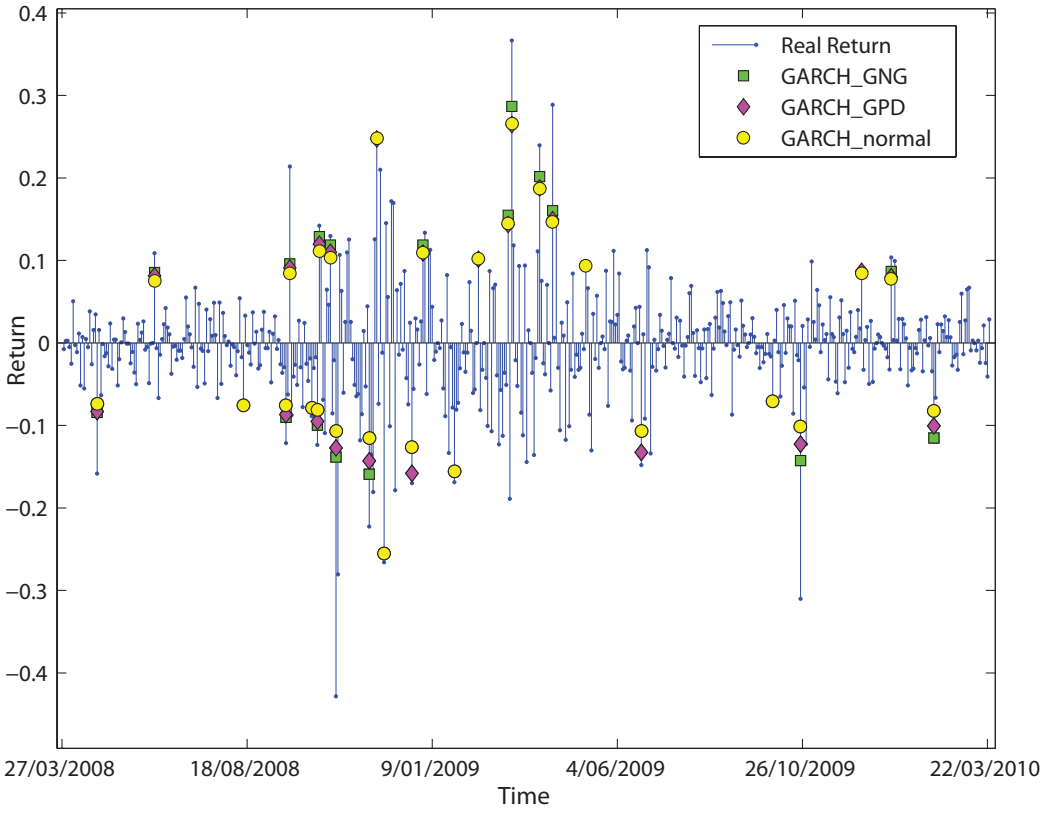

Figure 6.14: Violations of quantiles (99% & 1%) forecasting for PETC. The vertical lines are the actual returns of PETC and different symbols on the lines correspond to violation points based on different models as: square for GARCH-GNG, diamond for GARCH-GPD and circle for GARCH-normal.

tions corresponds to overestimation of the VaR. Figure 6.14 plots the violations of the PETC for $q \in (0.005, 0.01, 0.99, 0.995)$ using GARCH-GNG, GARCH-GPD and GARCH-normal during the current ongoing financial crisis time. It is clear that the violation rate is higher for GARCH-normal than GARCH-GNG and GARCH-GPD models since the extreme value type models are a better approximation of tail distributions and normal distribution generally under estimates the tail of financial series. Because our developed GNG model can estimate the threshold more objectively and account for the uncertainty associated with threshold choice, it returns a generally wider uncertainty interval compare to GPD approach. Figure 6.15 also reports the violations for the associated 95% confidence interval of these forecasted quantiles based on GARCH-GNG and GARCH-GPD models. Consistently the GARCH-GNG model has less violations than GARCH-GPD for the associated uncertainty interval of these quantile predictions.

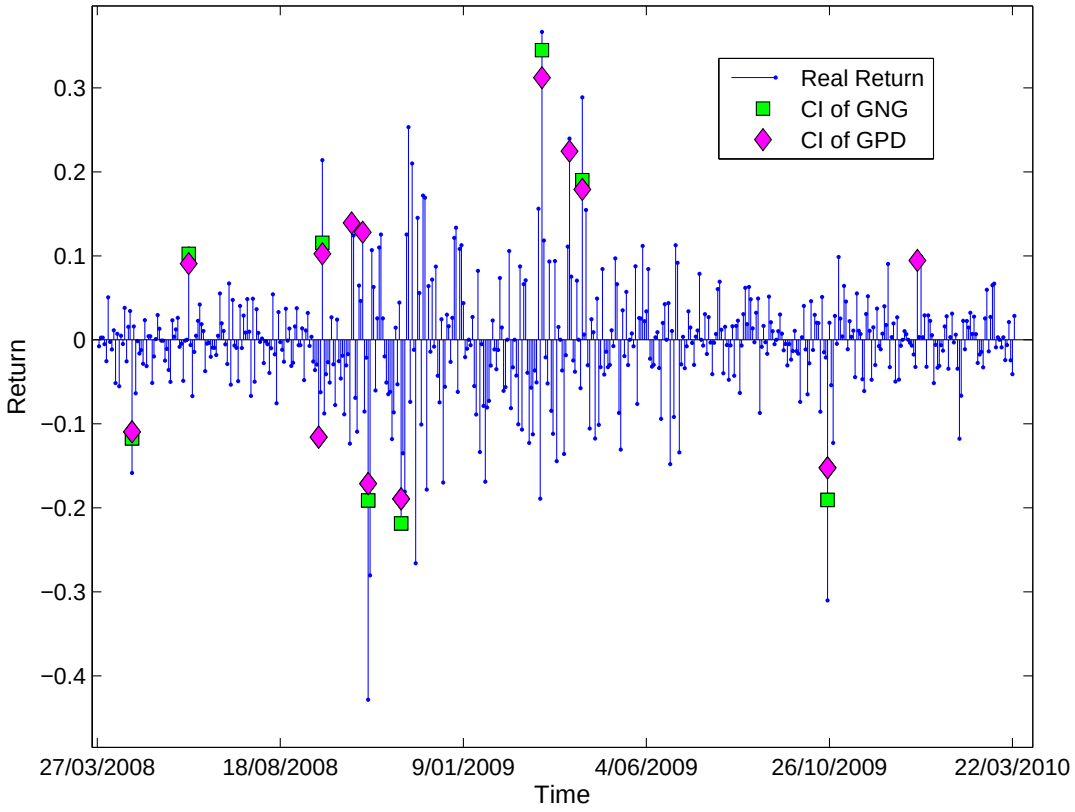

Figure 6.15: Violations of 95% CI for quantile (99% & 1%) forecasting for PETC. The vertical lines are the actual returns of PETC and different symbols on the lines correspond to violation points for uncertainty interval of quantile forecasting based on different models as: square for GARCH-GNG and diamond for GARCH-GPD.

The full results of violations for the whole forecasting period and four applications are shown in Table 6.3. A binomial test can be developed to test the success of these quantile estimation methods based on the number of violations. At each time point t, if we use a indicator $I_t = 1$ represents that a violation occurs and $I_t = 0$ represents that a violation

does not occur, the indicator follows a Bernoulli distribution as:

For upper quantiles,

$$I_t = \mathbf{1}_{\left(\hat{R}^q_{t+1} < R^q_t\right)} \sim Bernoulli(1-q)$$

For lower quantiles,

$$I_t = \mathbf{1}_{\left(\hat{R}^q_{t+1} > R^q_t\right)} \sim Bernoulli(q)$$

Therefore the total number of violations during the forecasting period T, containing n trading days, follows a binomial distribution as

$$\sum_{t=1}^{T} I_t \sim Binomial(n, p)$$

where, p is prescribed tail probability, $p = 1 - q$ for upper tail and $p = q$ for lower tail corresponding to a quantile R_q.

A two-sided binomial test is conducted under the null hypothesis that a method correctly estimates the conditional quantiles, or say that the estimated number of violations equals to theoretically expected number of violations. The alternative hypothesis is that the method estimates too few or too many violations. The corresponding p-value for the two-sided binomial test are reported in Table 6.3. A p-value less than 0.05 indicates the evidence against null hypothesis and a larger p-value indicates the evidence supporting the null hypothesis.

Both our model GNG and GPD have the flexibility in explaining the tail behavior whatever the underlying tail is heavy or not. However, normal and student t distribution would produce a biased quantile estimation when the underlying tail is not normal or as heavy as t distribution. As mentioned in Section 6.4.3 that normal or t distribution is more likely to under-estimate or over-estimate the tail of financial series, the variability is therefore high in extremal quantile predictions.

We highlight the best approach for all quantiles using the bold font in Table 6.3, e.g. GARCH-GNG is the best approach for JPM at quantile *0.005* with 13 violations and a p-value of *0.*7. In examining VaR, the violations for quantile at $q \in (0.005, 0.01, 0.99, 0.995)$ shows that our model GARCH-GNG is generally the best model for these applications and there is no case that our approach failed (rejection of the null hypothesis). GARCH-GPD is the second best approach and lots of failures for GARCH-normal and GARCH-t.

The results for relatively less extremal quantiles ($q \in (0.03, 0.05, 0.95, 0.97)$) are mixed, GARCH-GNG has a similar performance as GARCH-GPD and GARCH-normal. It is not surprising that the GARCH-GPD model performs well and similar to the proposed GARCH-GNG model, as it is a similarly flexible model in the tails. The performance of

Table 6.3: The comparison results in counting violations of quantile forecasting: Expected number of violations and number of violations obtained using GARCH-GNG (our approach), GARCH-GPD, GARCH-normal (GARCH with normally distributed innovations) and GARCH-t (GARCH with t distributed innovations). p-values for a binomial test are given in brackets and the length of test is *3000* for all applications. The table also gives the number of violations for the uncertainty interval (CI indicates confidence interval for MLE and credible interval for Bayesian.

	JPM		CITI		PETC		S&P100	
0.005 Quantile								
Expected	15		15		15		15	
GARCH-Mixture	***13***	***(0.70)***	***17***	***(0.70)***	***16***	***(0.90)***	***14***	***(0.90)***
GARCH-GPD	20	(0.24)	21	(0.15)	22	(0.09)	22	(0.09)
GARCH-Normal	36	(0.00)	34	(0.00)	33	(0.00)	45	(0.00)
GARCH-t	9	(0.15)	10	(0.24)	7	(0.05)	12	(0.52)
0.01 Quantile								
Expected	30		30		30		30	
GARCH-Mixture	***34***	***(0.52)***	***36***	***(0.31)***	***32***	***(0.78)***	***35***	***(0.41)***
GARCH-GPD	35	(0.41)	38	(0.17)	40	(0.08)	36	(0.31)
GARCH-Normal	47	(0.00)	54	(0.00)	50	(0.00)	69	(0.00)
GARCH-t	23	(0.23)	19	(0.05)	14	(0.00)	27	(0.65)
0.03 Quantile								
Expected	90		90		90		90	
GARCH-Mixture	***102***	***(0.22)***	118	(0.00)	123	(0.00)	107	(0.08)
GARCH-GPD	102	(0.22)	113	(0.02)	125	(0.00)	***106***	***(0.10)***
GARCH-Normal	105	(0.12)	***111***	***(0.03)***	***96***	***(0.56)***	129	(0.00)
GARCH-t	58	(0.00)	59	(0.00)	33	(0.00)	85	(0.63)
0.05 Quantile								
Expected	150		150		150		150	
GARCH-Mixture	170	(0.10)	181	(0.01)	184	(0.01)	173	(0.06)
GARCH-GPD	167	(0.17)	182	(0.01)	194	(0.00)	***169***	***(0.12)***
GARCH-Normal	***148***	***(0.90)***	***161***	***(0.38)***	***132***	***(0.14)***	178	(0.02)
GARCH-t	102	(0.00)	108	(0.00)	49	(0.00)	123	(0.03)
0.95 Quantile								
Expected	150		150		150		150	
GARCH-Mixture	175	(0.04)	156	(0.65)	***169***	***(0.12)***	168	(0.14)
GARCH-GPD	164	(0.26)	***151***	***(0.97)***	177	(0.03)	***161***	***(0.38)***
GARCH-Normal	***144***	***(0.65)***	136	(0.26)	***131***	***(0.12)***	129	(0.09)
GARCH-t	96	(0.00)	100	(0.00)	56	(0.00)	91	(0.00)
0.97 Quantile								
Expected	90		90		90		90	
GARCH-Mixture	97	(0.49)	***100***	***(0.31)***	95	(0.63)	106	(0.10)
GARCH-GPD	***90***	***(1.00)***	***100***	***(0.31)***	101	(0.26)	100	(0.31)
GARCH-Normal	96	(0.56)	99	(0.36)	***91***	***(0.96)***	***83***	***(0.49)***
GARCH-t	55	(0.00)	62	(0.00)	29	(0.00)	46	(0.00)
0.99 Quantile								
Expected	30		30		30		30	
GARCH-Mixture	***30***	***(1.00)***	***32***	***(0.78)***	***37***	***(0.23)***	***33***	***(0.65)***
GARCH-GPD	31	(0.93)	33	(0.65)	42	(0.03)	36	(0.31)
GARCH-Normal	40	(0.08)	45	(0.01)	57	(0.00)	34	(0.52)
GARCH-t	22	(0.17)	15	(0.01)	7	(0.00)	9	(0.00)
0.995 Quantile								
Expected	15		15		15		15	
GARCH-Mixture	***19***	***(0.36)***	14	(0.90)	***15***	***(1.00)***	***15***	***(1.00)***
GARCH-GPD	20	(0.24)	***15***	***(1.00)***	20	(0.24)	20	(0.24)
GARCH-Normal	31	(0.00)	30	(0.00)	42	(0.00)	16	(0.90)
GARCH-t	9	(0.15)	8	(0.09)	4	(0.01)	3	(0.00)

the GARCH-normal was a little bit of a surprise in some sense. However, this is likely caused by the normal distribution providing a good approximation out to these relatively low tail probabilities. This area in the range of support is typically more uncertainty for the GNG and GPD models due to the extra threshold uncertainty which substantially impacts the estimates close to the threshold.

In summary, for very small upper and lower tail probabilities we have shown that our proposed GARCH-GNG model outperforms the GARCH-normal, GARCH-t and GARCH-GPD models which are currently commonly used in the literature in VaR.

6.6 An Alternative Two Stage Model with Realized Volatility

6.6.1 Method

As the name, "realized" volatility is an observed volatility estimate rather than latent estimates using cumulative variance of the intra-day returns using high-frequency intra-day return data. Under suitable conditions, realized volatility is an unbiased and highly efficient estimator of return volatility as suggested by the theory of quadratic variation, as discussed in Andersen, Bollerslev, Diebold, and Labys (2001). We therefore consider a VaR estimate using realized volatility in the two stage model as an alternative when the case that truly high frequency data available to obtaining realized volatilities. Specifically, we standardize the return sequence by realized volatility in the first stage and model the innovation in the second stage with GNG mixture model. The procedure of VaR with realized volatility (assuming it has been calculated) is as the following:

1. Assume that $R_t = rv_t x_t$, here rv_t is the realized volatility at day t and x_t is the innovation term. Therefore the sequence of innovation can be simply calculated as $x_t = R_t/rv_t$.

2. Fit the proposed GNG mixture model to $\{x_t\}$ (the standardized innovation sequence) as described in Chapter 5.2.2. The upper tail of the mixture model represent short rises and the lower tail represents short falls.

Although the realized volatility is an observed volatility and model free, we need a model in forecasting the volatility for the next period in order to predict VaR. For simplicity, we assume a AR(1) process of the $\log(rv_t^2)$ as:

$$\log(rv_t^2) = \gamma_0 + \gamma_1 \log(rv_{t-1}^2) + \epsilon_t$$

We only use historical information from $n = 1000$ days in forecasting realized volatility rv_t and forecast the Rv_t at time $t-1$ according to the estimated model. The detailed scheme is as follows at time t:

1. With return sequence $\{R_{t-n-1}, ..., R_{t-1}\}$ and realized volatility $\{rv_{t-n-1}, ..., rv_{t-1}\}$, calculate the standardized innovation sequence $\{x_{t-n-1}, ..., x_{t-1}\}$.

2. Fit AR(1) model on $rv_{t-n-1}, ..., rv_{t-1}$ and obtain the one-step forecasting of the realized volatility $\widehat{rv_t}$ based on the fitted AR(1) model.

3. Apply the GNG model on the standardized innovation sequence $x_{t-n-1}, ..., x_{t-1}$, and forecast the expected quantile $\hat{x_t^q}$ based on the predictive posterior quantile distributions.

4. The forecasted return quantile can be calculated as $\hat{R}_t^q = \hat{rv}_t \hat{x_t^q}$.

Repeat all the steps for each time point to obtain a sequence of conditional forecasted return quantiles $\{\hat{R}_t^q\}$ for the forecasting period. Again these forecasted return quantiles are termed "conditional" as they are conditional on the variance being assumed known, using the estimates from the realized volatilities.

6.6.2 Empirical Results

Since we did not have access to the high frequency data for S&P100 and Citigroup which is required to calculate the realized volatility, we use different data set. The data used in this section is kindly provided by Scharth and Medeiros (2006), including realized volatility and returns of 14 stocks traded on the New York Stock Exchange and two traded on the NASDAQ as: Alcoa, American International Group, Boeing, Caterpillar, General Electric, General Motors, Hewlett Packard, IBM, Intel, Johnson and Johnson, Coca-Cola, Merck, Pfizer, Walmart and Exxon. The study period is from January 3, 1994 to December 31, 2003. Trading days with abnormally small trading volumes were excluded, leaving a total of 2539 daily observations. Unfortunately, we cannot examine the VaR for current ongoing crisis and only show the results for IBM for the period of 9.11 crisis in 2001 for brevity.

From Figure 6.16, the extreme quantile forecasted with realized volatilities for IBM shows higher variability compare to the quantile forecasted by GARCH-GPD model as the realized volatilities are generally more volatile than the latent volatilities (such as estimated by GARCH type models). The realized volatility adjusts faster to the market risk and therefore induce the VaR forecasted also change faster according to the market

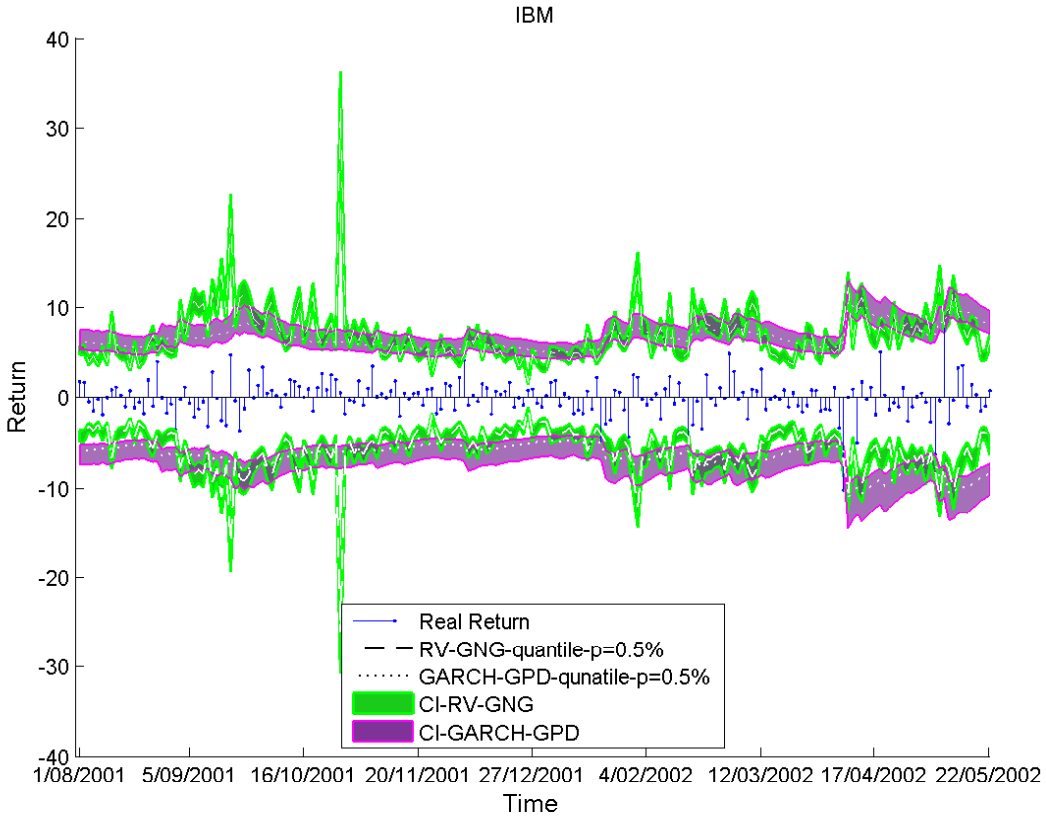

Figure 6.16: Conditional return quantile forecasting of IBM with realized volatilities. The dashed line is the conditional quantiles based on RV-GNG model and the interval in green is the uncertainty interval associated. The dotted line denotes the conditional qunatile forecasted based on the GARCH-GPD model and the interval in purple represents the uncertainty interval(CI) with it.

risk. Generally, the extreme quantiles forecasted in the crisis period (or see higher market risk period) will be further out from them forecasted with the latent volatilities and in reverse in the low risk period. Also, the standardized innovation term using realized volatilities is not as heavy tailed as the innovations using latent volatilities. This results in a narrower uncertainty interval compare to the two stage model with latent volatility.

6.7 Summary

In this paper, we proposed an approach for forecasting the VaR in combining the classical GARCH model and a new GPD mixture model. The proposed model overcomes the difficulty in threshold selection in traditional approaches and result in a more objective estimate of VaR. Further, the threshold is an explicit parameter of the model to be estimated, and therefore the uncertainty associated with the threshold selection is accounted for in forecasting the VaR. The mixture model for the GARCH innovations gives a flexible asymmetric distribution, with the GPD for capturing the gains and losses tail behaviours and more applicable with financial time series. We applied the model in forecasting the VaR for the Citigroup and S&P100 for two financial crisis period, 2001 and 2008, to demonstrate the advantages of the model for estimating the VaR. We compare our develop model with other methods in forecasting VaR by examining the violations of predictions. The result shows that for very small upper and lower tail probabilities, our proposed GARCH-GNG model outperforms the GARCH-normal, GARCH-t and GARCH-GPD models which are currently commonly used in the literature in VaR.

We also use alternative volatility realized volatility with proposed GNG model in estimate and forecasting VaR to see the differences from the GARCH-GNG model. The empirical results show that realized volatility based method tend to give a faster changing VaR and a narrower uncertainty interval associated compare with the method using latent volatilities because of a heavier volatility distribution and a thinner innovation distribution with it.

A natural extension/adaption of the proposed GARCH-GNG model in VaR is to consider an alternative distribution for the main mode of density, as discusses in Chapter 5 for those applications which is likely asymmetric feature around mode. A more flexible approach may be to adopt or extend the approach of Tancredi, Anderson, and O'Hagan (2006) to allow an unknown number of uniform distributions for the main mode of the density between the thresholds. This thesis mainly focuss on the extreme value modelling in VaR and only a simple method used in forecasting realized volatility in VaR. When truly high-frequency intra-day return is available, a more complex method can be considered in modelling realized volatility.

Chapter 7

Conclusion

7.1 Conclusion of Thesis

This thesis develops novel extreme value modelling in both neonatal and financial studies in risk analysis, and apply the stochastic volatility modelling with a particle filter in neonatal study. Econometric volatility models and time series methods have been combined with extreme value modelling to deal with known issues in applying extreme value models, such as dependence and threshold selection. The results presented in the thesis support following conclusions.

Chapter 3 build a statistical model to examine the variability of physiological measurements of preterm infants. Instabilities of preterm infants, resulting from their underdeveloped biological systems, stimulate the interest in studying the additional information contained in the underlying variability of physiological measurements on the patients to better understand and potentially classify their healthy status. The stochastic volatility model (SVM) is incorporated into an online sequential estimation framework, namely the ASIR particle filter, to model the conditional volatility of oxygen concentration of preterm infants. The empirical results show the SVM provides a reliable estimate of instantaneous latent volatility of medical measurements of preterm infants. In the thesis, we also apply extreme modelling with generalized pareto distribution in describing the tail behaviour of the changes of oxygen concentration level to indicate the potential risk of instability of preterm infants. The dependence of extremes is treated by the stochastic volatility clusters and the GPD is applied to the independent innovation term, standardized by the SVM volatility estimator, in estimating extreme quantiles.

In Chapter 4, we have constructed a new extreme value model GEV-GARCH for dynamic estimation of extremal quantiles and conditional variances. To capture the temporal

dependence typically observed in financial applications, that of volatility clustering, the scale and location parameters of the generalised extreme value distribution (GEV) have been defined to follow a conditional autoregressive heteroscedastic structure. Both the simulation study and real data application have been used to demonstrate the empirical performance of the GEV-GARCH in capturing the dynamic of conditional variance of extremes and to model the tail behaviour of the underlying variables.

In the comparison with other methods, GARCH (with normal innovation distribution) and two stage GARCH-GPD of McNeil and Frey (2000), the developed GEV-GARCH is shown to outperform the standard GARCH (with normal innovation distribution) for estimating the tail quantiles, due to flexibility of the extreme value distribution for capturing the heavy tail behaviour typically observed in financial times series. The GEV-GARCH provides similar performance to the two stage approach where the generalised Pareto distribution is used to describe the tail of the innovation distribution, but has an important benefit in that it is a one stage procedure, thus making parameter estimation and accounting for the uncertainties more straightforward.

Further simulation and empirical studies demonstrate the model identification and parameter estimation complications from permitting a time varying shape parameter with a similar GARCH structure, as proposed by Bali and Weinbaum (2007). The shape parameter is known to be challenging to estimate requiring very large sample sizes, see Coles (2001). The results in the thesis also suggests it is generally not realistic to have a varying shape for financial time series with even moderate sample sizes (of order 3000 days if high-frequency data considered too small) to have a tail behaviour that changes according to a GARCH-type structure.

Extreme value theory based models have been widely used in financial applications when assessing financial risk as they supply a statistically justifiable and flexible method for extrapolating tail distributions such as VaR. Chapter 5 shows that the commonly used excess over threshold - GPD model has difficulty in the threshold selection. The existing method of the mixture models, proposed in the literature, with the upper tail of the population distribution modeled using a GPD can solve the problem by automatic estimate threshold and therefore avoids subjective decision in threshold selection. However, these mixture models cannot be generally applied to financial data as often both gains and losses tails are both heavy tailed. In this thesis, we propose a two tail GPD mixture model which allows asymmetry of upper and lower tails and estimates threshold explicitly.

As the threshold is an explicit parameter of the model to be estimated, the subjective threshold choice can be avoided. Additionally, the proposed mixture model is able to account for the uncertainty associated with the threshold choice in latter inferences

about tail quantities like VaR.

A simulation study has been used to show the performance of a Bayesian inference approach for fitting the new GPD mixture model, and is shown to peform well when compared to robust estimation confirms approach in existing literature. The mixture model was also applied to various different population distributions (symmetric and asymmetric) and was shown to provide good approximations to various high quantiles, which indicate the general applicability of the proposed modelling approach.

As the threshold is estimated as part of the inference process, the model fitting (including threshold choice) is easily automated for large scale application to multiple time series which is very challenging for the traditional approaches to GPD threshold choice. The proposed two tail mixture model is very flexible in terms of tail features and is more applicable than the one tail mixture models.

Besides threshold choice, the dependence need to be treat in extreme modelling in estimating Value-at-Risk (VaR). In Chapter 6, the research therefore considers the two stage method by McNeil and Frey (2000) when applying proposed two GPD mixture model to imitate VaR. More specifically, we used proposed two tail GPD mixture model on the standard innovation term in the second stage instead of classical fixed-threshold GPD. This method of estimating VaR can overcome the difficulty of describing the dependence of extreme returns (from volatility clustering) and the challenge of threshold selection in traditional GPD applications. More importantly, it is able to account for the uncertainty associated with the threshold choice, as the threshold is also an parameter of the model to be estimated. The proposed VaR method is implemented in forecasting the VaR for both market index and individual stock returns during the financial crisis periods (2001 and current). The automatic threshold selection procedure allows the different threshold value chosen conditional on time windows, which making the forecasting easier compared to the classical GPD model. Particularly, we compare our develop model with other methods in forecasting VaR by examining the violations of predictions. The result shows that for very small upper and lower tail probabilities, such as 1% or 0.5% tail probability used in VaR, our proposed GARCH-GNG model outperforms the GARCH-normal, GARCH-t and GARCH-GPD models which are currently commonly used in the literature in VaR. Additionally, we consider a two stage VaR in using realized volatility in the first stage and proposed two GPD mixture model in the second stage in forecasting VaR. The empirical results show that realized volatility based method tended to give a faster changing VaR and a narrower uncertainty interval compared with the method using latent volatilities because of a heavier volatility distribution and a thinner innovation distribution.

7.2 Discussion of Future Research

The neonatal research in this thesis forms part of the initial stages of neonatal research project, which is carried by Christchurch Women's Hospital and University of Canterbury. Chapter 3 provides insight into the stochastic volatility modeling approach and how it can be used to estimate the high quantiles (which are of interest for determining normal ranges) for a single patient only. Future research will need to consider how information across multiple patients can be pooled to provide physicians with estimates of normal ranges of the volatility at difference gestational within the population.

In Chapter 4, we discussed the model identification issue related to a time varying shape parameter and suggest further research could investigate a smoothly varying (Davison and Ramesh 2000b; Conigliani and Tancredi 2005), or structural break type model for the shape parameter if there is believed to be some form of change in market trading behaviour.

The two tail GPD mixture model proposed in Chapter 5 considers the normal distribution for the main mode of the population distribution for financial applications, due to their inherent unimodal, approximately symmetric and quadratic shape around the mode. However, for applications where an asymmetric mode is expected, alternative distributions for the bulk should be considered, e.g. a Weibull or gamma distribution. The choice of the distribution used to capture the main mode of the distribution in the mixture model (normal considered in this paper) was shown in the simulation study to affect the performance of the model in capturing the quantiles. Alternatively, a more flexible distribution would be a natural extension for the bulk distribution, such as a mixture of uniform distributions as the approach of Tancredi, Anderson, and O'Hagan (2006) or non-parametric smooth method to overcome the asymmetry and multimodal. As a consequence, a more complex model of this form will be required with the associated extra estimation and computation challenges.

In Chapter 6, we applied the proposed two GPD tail mixture model to estimating VaR, in combination with the two stage model suggested by McNeil and Frey (2000), which accounts for the volatility clustering in the first stage. Although the automated estimation of threshold accounts for the uncertainty associated with it's estimation uncertainty, the two stage model still suffers in not accounting for the uncertainty of the volatility estimation in the first stage in latter inference. Further, research is needed to consider either a one stage approach similar to that we developed in Chapter 4, or alternatively a full uncertainty analysis which can account for all uncertainties in the VaR inferences.

Appendix A

Figures

A.1 Density Plot of GNG Fitted to General Distributions

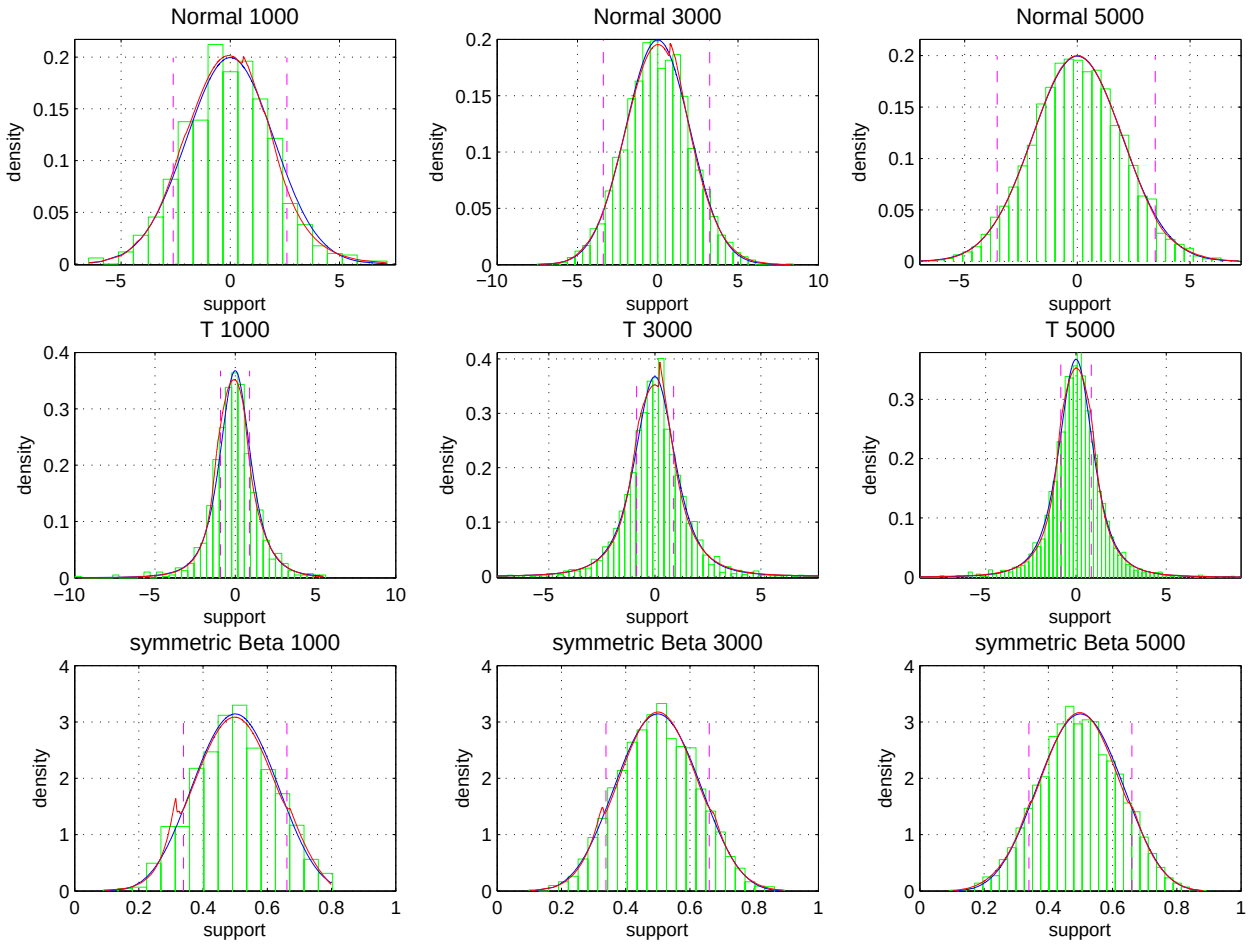

Figure A.1: Density of GNG on symmetric distributions. Data are from simulation study II in Section 5.5. The figure only show the density of one sample for each distribution for each different sample size. The red line is fitted density by GNG model and the blue line is the true density under the true distribution functions.

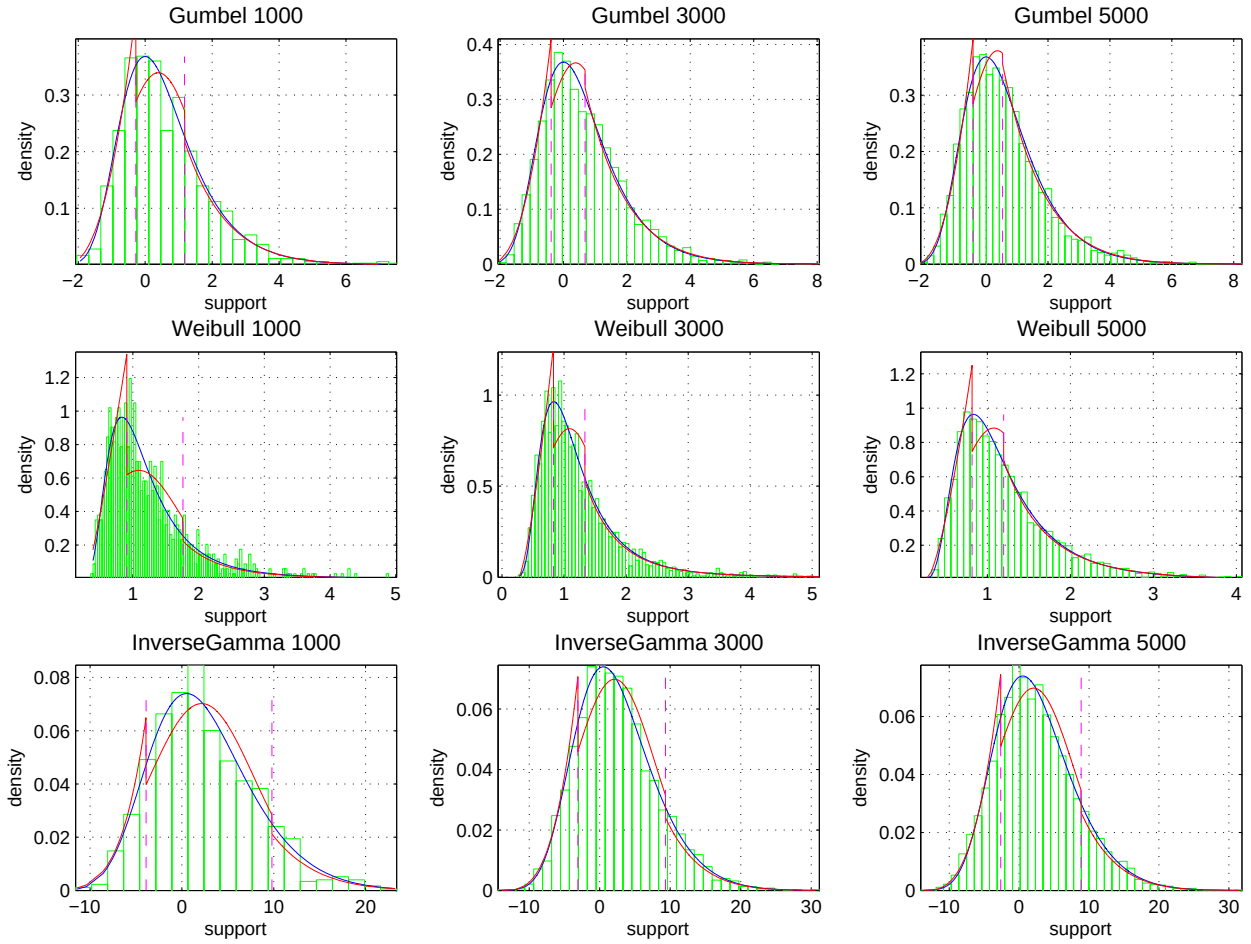

Figure A.2: Density of GNG on asymmetric distributions. Data are from simulation study II in Section 5.5. The figure only show the density of one sample for each distribution for each different sample size. The red line is fitted density by GNG model and the blue line is the true density under the true distribution functions.

Appendix B

Algorithm of MCMC for Two Tail GPD Mixture Model

This appendix gives the summary of Markov chain Monte Carlo algorithm for the two tail GPD mixtrue model builded in Chapter 5, see detailed information of posterior and prior in Section 5.3. Posteriors sampling is through Metropolis-Hastings (M-H) algorithm as it has the advantage of being free of functional form when the posterior distribution function is not a proper probability function. In particular, random walk M-H is used for the method of sampling.

B.1 Summary of MCMC Algorithm

Suppose the two tail GPD mixture model parameter set is $\theta^j = (m^j, s^j, u_r^j, \xi_r^j, \sigma_r^j, u_l^j, \xi_l^j, \sigma_l^j)$ at iteration j of the MCMC chain. The parameter is sampled in the order of importance of tail distribution as $(\xi_r^j, \sigma_r^j, u_r^j, \xi_l^j, \sigma_l^j, u_l^j, m^j, s^j)$. The algorithm of jth iteration of the chain in updating θ^j from θ^{j-1} is as following:

1. let $\theta^* = \theta^{j-1}$. We will use p denote the posterior probability in the following.

2. Sampling ξ_r
   Draw a new value, ξ_r^*, using a truncated normal distribution (proposal distribution) $N(\xi_r^{j-1}, V_{\xi_r})I(l\xi_r, \infty)$, where V_{ξ_r} is pre-determined based on the curvature of the conditional posterior mode. The new draw must line within the support region of $\xi_r \in (l\xi_r, \infty)$ with the lower boundary $l\xi_r = -\sigma_r^{j-1}/(max\{x\} - u_r^{j-1})$ to satisfy general condition of GPD excess x: $1 + \xi\frac{x-u}{\sigma} > 0$. The acceptance probability α_{ξ_r}

is:

$$\alpha_{\xi_r} = \min\left\{1, \frac{p(\theta^*|x)}{p(\tilde{\theta}|x)} \frac{\Phi\left(\frac{\xi_r^{j-1} - l\xi_r}{\sqrt{V_{\xi_r}}}\right)}{\Phi\left(\frac{\xi_r^* - l\xi_r}{\sqrt{V_{\xi_r}}}\right)}\right\}$$

for $\theta^* = (m^{j-1}, s^{j-1}, u_r^{j-1}, \xi_r^*, \sigma_r^{j-1}, u_l^{j-1}, \xi_l^{j-1}, \sigma_l^{j-1})$, $\tilde{\theta} = \theta^{j-1}$ and Φ denotes standard normal cumulative distribution. Draw a random value β from uniform $U(0,1)$. Let $\tilde{\theta} = \theta^*$ if $\beta < \alpha_{\xi_r}$, $\tilde{\theta} = \tilde{\theta}$ else.

3. Sampling σ_r

   Draw σ_r^* according to ξ_r^j of $\tilde{\theta}$.

   If $\xi_r^j \geq 0$, draw σ_r^* from Gamma distribution $G((\sigma_r^{j-1})^2/V_{\sigma_r}, V_{\sigma_r}/\sigma_r^{j-1})$. The acceptance probability α_{σ_r} is:

$$\alpha_{\sigma_r} = \min\left\{1, \frac{p(\theta^*|x)}{p(\tilde{\theta}|x)} \frac{G\left(\sigma_r^{j-1}|(\sigma_r^*)^2/V_{\sigma_r}, V_{\sigma_r}/\sigma_r^*\right)}{G\left(\sigma_r^*|(\sigma_r^{j-1})^2/V_{\sigma_r}, V_{\sigma_r}/\sigma_r^{j-1}\right)}\right\}$$

   If $\xi_r^j < 0$, draw σ_r^* from truncated normal distribution $N(\sigma_r^{j-1}, V_{\sigma_r})I(l\sigma_r, \infty)$, where lower boundary $l\sigma_r = -\xi_r^j(max\{x\} - u_r^{j-1})$ to satisfy general condition of GPD excess x: $1 + \xi\frac{x-u}{\sigma} > 0$. The acceptance probability α_{σ_r} is:

$$\alpha_{\sigma_r} = \min\left\{1, \frac{p(\theta^*|x)}{p(\tilde{\theta}|x)} \frac{\Phi\left(\frac{\sigma_r^{j-1} - l\sigma_r}{\sqrt{V_{\sigma_r}}}\right)}{\Phi\left(\frac{\sigma_r^* - l\sigma_r}{\sqrt{V_{\sigma_r}}}\right)}\right\}$$

   for $\theta^* = (m^{j-1}, s^{j-1}, u_r^{j-1}, \xi_r^j, \sigma_r^*, u_l^{j-1}, \xi_l^{j-1}, \sigma_l^{j-1})$, and Φ denotes standard normal cumulative distribution.

   Draw a random value β from uniform $U(0,1)$. Let $\tilde{\theta} = \theta^*$ if $\beta < \alpha_{\sigma_r}$, $\tilde{\theta} = \tilde{\theta}$ else.

4. Sampling u_r

   Draw a new value, u_r^*, using a truncated normal distribution $N(u_r^{j-1}, V_{u_r})I(lu_r, max\{x\})$, where $lu_r = min\{x\}$ if $\xi_r^j \geq 0$ and $lu_r = \max\{min\{x\}, max\{x\} + \sigma_r^j/\xi_r^j\}$ if $\xi_r^j < 0$.

   Then the acceptance probability α_{u_r} is:

$$\alpha_{u_r} = \min\left\{1, \frac{p(\theta^*|x)}{p(\tilde{\theta}|x)} \frac{\Phi\left(\frac{max\{x\} - u_r^{j-1}}{\sqrt{V_{u_r}}} - \frac{lu_r - u_r^{j-1}}{\sqrt{V_{u_r}}}\right)}{\Phi\left(\frac{max\{x\} - u_r^*}{\sqrt{V_{u_r}}} - \frac{lu_r - u_r^*}{\sqrt{V_{u_r}}}\right)}\right\}$$

   for $\theta^* = (m^{j-1}, s^{j-1}, u_r^*, \xi_r^j, \sigma_r^j, u_l^{j-1}, \xi_l^{j-1}, \sigma_l^{j-1})$, and Φ denotes standard normal

cumulative distribution.
Draw a random value β from uniform $U(0,1)$. Let $\tilde{\theta} = \theta^*$ if $\beta < \alpha_{u_r}$, $\tilde{\theta} = \tilde{\theta}$ else.

5. Sampling ξ_l
Draw a new value, ξ_l^*, using a truncated normal distribution (proposal distribution) $N(\xi_l^{j-1}, V_{\xi_l})I(l\xi_l, \infty)$, where V_{ξ_l} is pre-determined based on the curvature of the conditional posterior mode. The new draw must line within the support region of $\xi_l \in (l\xi_l, \infty)$ with the lower boundary $l\xi_l = -\sigma_l^{j-1}/(-min\{x\} + u_l^{j-1})$ to satisfy general condition of GPD. The acceptance probability α_{ξ_l} is:

$$\alpha_{\xi_l} = \min\left\{1, \frac{p(\theta^*|x)}{p(\tilde{\theta}|x)} \frac{\Phi\left(\frac{\xi_l^{j-1} - l\xi_l}{\sqrt{V_{\xi_l}}}\right)}{\Phi\left(\frac{\xi_l^* - l\xi_l}{\sqrt{V_{\xi_l}}}\right)}\right\}$$

for $\theta^* = (m^{j-1}, s^{j-1}, u_r^j, \xi_r^j, \sigma_r^j, u_l^{j-1}, \xi_l^*, \sigma_l^{j-1})$, and Φ denotes standard normal cumulative distribution. Draw a random value β from uniform $U(0,1)$. Let $\tilde{\theta} = \theta^*$ if $\beta < \alpha_{\xi_l}$, $\tilde{\theta} = \tilde{\theta}$ else.

6. Sampling σ_l
Draw σ_l^* according to ξ_l^j of $\tilde{\theta}$.
If $\xi_l^j \geq 0$, draw σ_l^* from Gamma distribution $G((\sigma_l^{j-1})^2/V_{\sigma_l}, V_{\sigma_l}/\sigma_l^{j-1})$. The acceptance probability α_{σ_l} is:

$$\alpha_{\sigma_l} = \min\left\{1, \frac{p(\theta^*|x)}{p(\tilde{\theta}|x)} \frac{G\left(\sigma_l^{j-1}|(\sigma_l^*)^2/V_{\sigma_l}, V_{\sigma_l}/\sigma_l^*\right)}{G\left(\sigma_l^*|(\sigma_l^{j-1})^2/V_{\sigma_l}, V_{\sigma_l}/\sigma_l^{j-1}\right)}\right\}$$

If $\xi_l^j < 0$, draw σ_l^* from truncated normal distribution $N(\sigma_l^{j-1}, V_{\sigma_l})I(l\sigma_l, \infty)$, where lower boundary $l\sigma_l = -\xi_l^j(-min\{x\} + u_l^{j-1})$ to satisfy general condition of GPD excess x: $1 + \xi\frac{x-u}{\sigma} > 0$. The acceptance probability α_{σ_l} is:

$$\alpha_{\sigma_l} = \min\left\{1, \frac{p(\theta^*|x)}{p(\tilde{\theta}|x)} \frac{\Phi\left(\frac{\sigma_l^{j-1} - l\sigma_l}{\sqrt{V_{\sigma_l}}}\right)}{\Phi\left(\frac{\sigma_l^* - l\sigma_l}{\sqrt{V_{\sigma_l}}}\right)}\right\}$$

for $\theta^* = (m^{j-1}, s^{j-1}, u_r^j, \xi_r^j, \sigma_r^j, u_l^{j-1}, \xi_l^j, \sigma_l^*)$, and Φ denotes standard normal cumulative distribution.
Draw a random value β from uniform $U(0,1)$. Let $\tilde{\theta} = \theta^*$ if $\beta < \alpha_{\sigma_l}$, $\tilde{\theta} = \tilde{\theta}$ else.

7. Sampling u_l
Draw a new value, u_l^*, using a truncated normal distribution

$N(u_l^{j-1}, V_{u_l})I(min\{x\}, Uu_r)$, where Uu_r is the upper boundary and $Uu_l = min\{x\}$ if $\xi_l^j \geq 0$ and $Uu_l = \min\left\{max\{x\}, max\{-x\} + \sigma_l^j/\xi_l^j\right\}$ if $\xi_l^j < 0$.
Then the acceptance probability α_{u_l} is:

$$\alpha_{u_L} = \min\left\{1, \frac{p(\theta^*|x)}{p(\tilde{\theta}|x)} \frac{\Phi\left(\frac{Uu_l - u_L^{j-1}}{\sqrt{V_{u_L}}} - \frac{min\{x\} - u_L^{j-1}}{\sqrt{V_{u_L}}}\right)}{\Phi\left(\frac{Uu_l - u_L^*}{\sqrt{V_{u_L}}} - \frac{min\{x\} - u_L^*}{\sqrt{V_{u_L}}}\right)}\right\}$$

for $\theta^* = (m^{j-1}, s^{j-1}, u_r^j, \xi_r^j, \sigma_r^j, u_l^*, \xi_l^j, \sigma_l^j)$, and Φ denotes standard normal cumulative distribution.
Draw a random value β from uniform $U(0,1)$. Let $\tilde{\theta} = \theta^*$ if $\beta < \alpha_{u_l}$, $\tilde{\theta} = \tilde{\theta}$ else.

8. Sampling m and s
Draw m^* from normal distribution $N(m^{j-1}, V_m)$ and draw s^* from Gamma distribution $G((s^{j-1})^2/V_m, V_m/s^{j-1}))$. Assume m and s are independent. The acceptance probability $\alpha_{m,s}$ is:

$$\alpha_{m,s} = \min\left\{1, \frac{p(\theta^*|x)}{p(\tilde{\theta}|x)} \frac{\Phi\left(\frac{m^{j-1} - m^*}{V_m}\right)}{\Phi\left(\frac{m^* - m^{j-1}}{V_m}\right)} \frac{G(s^{j-1}|(s^*)^2/V_s, V_s/s^*)}{G(s^*|(s^{j-1})^2/V_s, V_s/s^{j-1})}\right\}$$

for $\theta^* = (m^*, s^*, u_r^j, \xi_r^j, \sigma_r^j, u_l^j, \xi_l^j, \sigma_l^j)$, where Φ and G denotes standard normal and Gamma cumulative distribution respectively.
Draw a random value β from uniform $U(0,1)$. Let $\theta^j = \theta^*$ if $\beta < \alpha_{m,s}$, $\theta^j = \tilde{\theta}$ else.

9. Let $j = j + 1$ and Go back to step 1. Algorithm is iterated, from step 1 to 8, a large number of times.

B.2 Implementations of MCMC

In order to get the variance (V_θ) of the parameter by approximation of the curvature of posterior distribution at its mode for each parameter in above MCMC algorithm, we first run a chain without fixing V_θ. V_θ is updated at each iteration of the pre-chain until it converged to a constant. Then a few chain with different starting value are used to check the convergence of the chain and for tuning V_θ in order to reach all states of posterior distribution with a reasonable overall acceptance rate (Gelman, Carlin, Stern, and Rubin 2004).

Only the converged part of the chain is regarded as an approximate sample from the posterior distributions. The convergence of the iterative simulation of posterior is assessed

by monitoring multiple simulated sequence with dispersed starting values as suggested by Gelman, Carlin, Stern, and Rubin (2004). We can estimate the marginal posterior variance by a weighted average of the between (B) and within (W) sequence variances for each of parameter estimator as $\hat{var}(\theta_i|x) = \frac{n-1}{n}W_{var} + \frac{1}{n}B_{var}$, here n indicates the length of simulation. This is commonly used to the classical variance estimate with cluster sampling. For a finite n, the within variance will under-estimate $var(\theta_i|x)$ because the individual chain has not reach the entire target distribution. The expectation of W approaches $var(\theta_i)$ as the limit $n \to \infty$. We assess the convergence by monitoring that the scale of the current posterior distribution for θ_i might be reduced if the simulation continue in the limit $n \to \infty$. The potential scale reduction is defined as the ratio of the marginal posterior variance and within variance as $\hat{R} = \sqrt{\hat{var}(\theta_i|x)/W}$, which should decline to 1 as the chain length goes to infinity. The second half of the chain is used as the posterior distribution and the estimated parameter is calculated as the mean of highest posterior density interval (HPD) of the posteriors.

Bibliography

Aït-Sahalia, Y., P. Mykland, and L. Zhang (2005). How often to sample a continues-time in the presence of the market microstructure noise. *The Review of Finacial Studies 18*, 351–416.

Andersen, T. G. and T. Bollerslev (1998). Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance 53*, 219–265.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics 61*, 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association 96*, 42–55.

Andersen, T. G., T. Bollerslev, and S. Lange (1999). Forecasting financial market volatility: Sample frequency vis-á-vis forecast horizon. *Journal of Empirical Finance 6*, 457–477.

Andersen, T. G., T. Bollerslev, and N. Meddahi (2005). Correcting the errors: Volatility forecast evaluation using high-frequency data and realized volatilities. *Econometrica 73*, 279–296.

Andreou, E. and E. Ghysels (2002). Rolling-sample volatility estimators: Some new theoretical, simulation and empirical results. *Journal of Business and Economic Statistics 20*, 363–376.

Bali, T. G. (2000). Testing the empirical performance of stochastic volatility models of the short term interest rate. *Journal of Financial and Quantitative Analysis 35*, 191–215.

Bali, T. G. (2003a). Disturbing extremal behavior of spot rate dynamics. *Journal of Empirical Finance 10*, 455–477.

Bali, T. G. (2003b). An extreme value approach to estimating volatility and value at risk. *Journal of Business 76*, 83–108.

Bali, T. G. and D. Weinbaum (2007). A conditional extreme value volatility estimator based on high–frequency returns. *Journal of Economic Dynamics and Control 31*, 361–397.

Barndorff-Nielsen, O. and N. Shephard (2002a). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society B 64*, 253–280.

Barndorff-Nielsen, O. E. and D. R. Cox (1994). *Inference and Asymptotics*. Chapman and Hall.

Barndorff-Nielsen, O. E. and N. Shephard (2002b). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society B 64*, 253–280.

Barndorff-Nielsen, O. E. and N. Shephard (2002c). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics 17*, 457–477.

Behrens, C. N., H. F. Lopes, and H. F. Gamerman (2004). Bayesian analysis of extrme events with threshold estimation. *Statistical Modelling 4*, 227–244.

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of Extremes: Theory and Applications*. Wiley Series in Probability and Statistics.

Beirlant, J., P. Vynckier, and J. L. Teugels (1996). Excess functions and estimation of the extreme-value index. *Bernoulli 2*, 293–318.

Beresford, M. W., H. Parry, and N. J. Shaw (2005). Twelve-month prospective study of oxygen saturation measurements among term and preterm infants. *Journal of Perinatology 25*, 30–32.

Bermudez, P. d. Z., M. A. Amaral Turkman, and K. F. Turkman (2002). A predictive approach to tail probability estimation. *Extremes 4*, 295–314.

Blair, B.J., P. S.-H. T. S. (2001). Forecasting s&p 100 volatility: The incremental information content of implied volatilities and high–frequency index returns. *Journal of Econometrics 105*, 5–26.

Bollen, N. and R. Whaley (2004). Does net buying pressure affect the shape of the implied volatility functions? *Journal of Finance 59*, 711–753.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*, 307–327.

Bollerslev, T., R. Y. Chou, and K. F. Kroner (1992). Arch modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics 52*, 5–59.

Buishand, T. (1993). *Statistics for the Environment*, Chapter Rainfall depth-duration-frequency curves: A problem of dependent extremes, pp. 183–197. Wiley.

Carnero, A., D. P. n. and E. Ruiz (2001). *Is Stochastic Volatility more Flexible than GARCH?* Universidad Carlos III de Madrid. Working Paper.

Chan, N. H., S. Deng, L. Peng, and Z. Xia (2007). Interval estimation of Value-at-Risk based on GARCH models with heavy–tailed innovations. *Journal of Econometrics 137*, 556–576.

Chen, T., J. Morris, and E. Martin (2005). Particle filters for state and parameter estimation in batch processes. *Journal of Process Control 15*, 665–673.

Christensen, B. J. and N. R. Prabhala (1998). The relation between implied volatility and realized volatility. *Journal of Financial Economics 50*, 125–150.

Coles, S. G. (2001). *An introduction to Statistical Modelling of Extreme Values*. Springer-Verlag.

Coles, S. G. and L. Pericchi (2003). Anticipating catastrophes through extreme value modelling. *Apllied Statistics 52*, 405–416.

Coles, S. G. and E. A. Powell (1996). Bayesian methods in extreme value modelling: A review and new developments. *International Statistical Review 64*, 119136.

Coles, S. G. and J. Tawn (1996). A Bayesian analysis of extreme rainfall data. *Applied Statistics 45*(4), 463–478.

Coles, S. G. and J. A. Tawn (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society B 53*, 377–392.

Coles, S. G. and J. A. Tawn (1994). Statistical methods for multivariate extremes: An application to structural design. *Applied Statistics 43*, 1–48.

Conigliani, C. and A. Tancredi (2005). Semi-parametric modelling for costs of health care technologies. *Statistics in Medicine 24*, 3171–3184.

Danielsson, J. and C. de Vries (1997). Beyond the sample: Extreme quantile and probability estimation. *Journal of Empirical Finance 4*, 241–257.

Davison, A. C. and N. I. Ramesh (2000a). Impact of urbanization on coastal wetland structure and function. *Journal of the Royal Statistical Society Series B 62*, 191–208.

Davison, A. C. and N. I. Ramesh (2000b). Local likelihood smoothing of sample extremes. *Journal of the Royal Statistical Society B 62*, 191–208.

Davison, A. C. and R. L. Smith (1990). Models for exceedence over high thresholds (with discussion). *Journal of the Royal Statistical Society 52*, 393–442.

Djurić, P. M., J. H. Kotecha, J. Zhang, T. G. Y. Huang, M. F. Bugallo, and J. Mj́guez (2003). Application of particle filtering to selected problem in communications. *Signal Processing Magazine September*, 19–38.

Doucet, A. (1998). On sequential Monte Carlo methods for Bayesian filtering. Technical report, Department of Engineering, University of Cambridge.

Duffie, D. and J. Pan (1997). An overview of Value-at-Risk. *Journal of Derivatives Spring*, 7–49.

Dupuis, D. J. (1998). Exceedances over high threshold: A guide to threshold selection. *Extremes 1*, 251–261.

Embrechts, P. (1999). *Extreme value theory in finance and insurance*. Zurich, Switzerland: Department of Mathematics, ETH, Swiss Federal Technical University.

Embrechts, P. (2000). Extreme value theory: Potentials and limitations as an integrated risk management tool. *Manuscript. Zurich, Switzerland: Department of Mathematics, ETH, Swiss Federal Technical University 20*, 287–303.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Springer–Verlag.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica 50*, 987–1008.

Engle, R. F. and V. K. Ng (1993). Measuring and testing the impact of news on volatility. *Journal of Finance 48*, 1749–1778.

Fearnhead, P. (2005). Using random Quasi-Monte-Carlo within particle filters with application to financial time series. *Journal of Computational and Graphical Statistics 14*, 751–769.

Ferro, C. A. T. and J. Segers (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society, Series B 65*(2), 545–556.

Fleming, J., C. Kirby, and B. Ostdiek (2003). The economic value of volatility timing using realized volatility. *Journal of Financial Economics 67*, 473–509.

Fleming, J., B. Ostdiek, and R. E. Whaley (1995). Predicting stock market volatility: a new measure. *Journal of Futures Markets 15*, 256–302.

Frances, P. H. and D. V. Dijk (1995). Forecasting stock market volatility using (non-linear) GARCH models. *Journal of Forecasting 15*, 229–235.

French, K. R., G. W. Schwert, and R. E. Stambaugh (1987). Expected stock returns and volatility. *Journal of Financial Economics 19*, 3–29.

Frigessi, A., O. Haug, and H. Rue (2002). A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes 5*, 219–235.

Galland, B. C., R. M. Hayman, B. J. Taylor, D. P. G. Bolton, R. M. Sayers, and S. M. Williams (2000). Factors affecting heart rate variability and heart rate responses to tilting in infants aged 1 and 3 months. *Pediatric Research 48*(3), 360–368.

Gamerman, D. (1997). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*. Chapman and Hall: London.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004). *Bayesian data analysis*. Chapman and Hall.

Gencay, R. and F. Selcuk (2004). Extreme value theory and Value-at-Risk: Relative performance in emerging markets. *International Journal of Forecasting 20*, 287–303.

Gençay, R., F. Selçuk, and A. Ulugülyağci (2003). High volatility, thick tails and extreme value theory in Value-at-Risk estimation. *Insurance: Mathematics and Economics 33*, 337–356.

Giot, P. and S. Laurent (2004). Modelling daily Value-at-Risk using realized volatility and ARCH type models. *Journal of Empirical Finance 11*(3), 379–398.

Glosten, L., R. Jagannathan, and D. Runkle (1992). On the relation between the expected value and volatility of nominal excess return on stocks. *Journal of Finance 46*, 1779–1801.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika 82*, 711–732.

Guillou, A. and P. Hall (2001). A diagnostic for selecting the threshold in extreme value theory. *Journal of the Royal Statistical Society B63*, 293–305.

Hample, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel (1986). *Robust statistics:The approach based on the Influence functions*. Wiley, New York.

Heffernan, J. E. and J. A. Tawn (2004). A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society B B66*, 1–34.

Hentschel, L. E. (1995). All in the family: Nesting symmetric and asymmetric garch models. *Journal of Financial Economics 39*, 71–104.

Hosking, J. and J. R. Wallis (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics 29*, 339–349.

Hsing, T. (1987). On the characterization of certain point processes. *Stochastic Process Application 26*, 297–316.

Hsing, T. (1993). Extremal index estimation for a weakly dependent stationary sequence. *The Annals of Statistics 21*, 2043–2071.

Hsing, T., J. Hüsler, and M. R. Leadbetter (1988). On the exceedance point process for a stationary sequence. *Probability. Theory Related Fields 78*, 97–112.

Jorion, P. (2000). *Value-at-Risk: The new benchmark for managing financial risk*. McGraw-Hill: New York.

Leadbetter, M. R. (1983). Extremes and local dependence of stationary sequences. *Z.Wahrscheinlichkeitstheorie verw. Gebiete 65*, 291–360.

Leadbetter, M. R. (1995). On high-level exceedance modeling and tail-inference. *J. Statist. Planng Inf. 45*, 247–260.

Leadbetter, M. R., I. Weissman, L. de Haan, and H. Rootzen (1989). On clustering of high values in statistically stationary series. *Technical Report 253. Center for Stochastic Processes, University of North Carolina, Chapel Hill*.

Liu, J. and M. West (2001). *Sequential Monte Carlo in Practice*, Chapter Combined parameter and state estimation in simulation-based filtering, pp. 197–223. Springer-Verlag, New York.

Longin, F. M. (2000). From value at risk to stress testing: the extreme value approach. *Journal of Banking and Finance 24*, 1097–1130.

Magni, P. and R. Bellazzi (2006). A stochastic model to assess the variability of blood glucose time series in diabetic patients self–monitoring. *IEEE Transactions on Biomedical Engineering 53*, 977–985.

Matthews, R. (1996). Far out forecasting. *The New Scientist 2051 12*, 36–40.

McAleer, M. and M. C. Medeiros (2008). Realized volatility: A review. *Econometric Reviews 27*, 10–45.

McNeil, A. and T. Saladin (1996). The econometrics of financial markets. *Journal of Empirical Finance 3*, 15–102.

McNeil, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin 27*, 117–137.

McNeil, A. J. (1998). *Calculating quantile risk measures for financial time series using extreme value theory*. Zurich, Switzerland: Department of Mathematics,ETH, Swiss Federal Technical University.

McNeil, A. J. (1999). *Extreme value theory for risk managers*. Zurich, Switzerland: Department of Mathematics,ETH, Swiss Federal Technical University.

McNeil, A. J. and R. Frey (2000). Estimation of tail-related risk measures for heteroscedastic financial time series an extreme value approach. *Journal of Empirical Finance 7*, 271–300.

Melino, A. T. and M. Stuart (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics 45*, 239–265.

Mendes, B. V. M. and H. F. Lopes (2004). Data driven estimates for mixtures. *Computational Statistics & Data Analysis 47*, 583–598.

Nelson, D. B. (1991). Conditional heteroscedasticity in asset returns: A new approach. *Econometrica 59*, 347–370.

Ng, A., N. Subhedar, R. A. Primhak, and N. j. Shaw (1998). Arterial oxygen saturation profiles in healthy preterm infants. *Archives of Disease in Childhood - Fetal and Neonatal Edition 79*, 64–66.

Oomen, R. C. A. (2002). Modelling realized variance when returns are serially correlated. *Upublished manuscript, Graduate School of Business School,University of Warwick*.

Oomen, R. C. A. (2005). Properties of bias-corrected realized variance under alternative sampling schemes. *Journal of Financial Econometrics 3*, 555–577.

Pauli, F. and S. G. Coles (2001). Penalized likelihood inference in extreme value theory. *Journal of Applied Statistics 28*, 547–560.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics 3*, 119–131.

Pitt, M. and N. Shephard (1999). Filtering via simulation: Auxiliary particle filters. *Journal of American Statistical Association 94*, 590–599.

Rachev, S. T., J. S. J. Hsu, B. S. Bagasheva, and F. J. Fabozzi (2008). *Bayesian Methods in Finance*. John Wiley and Sons, Inc., Hoboken, New Jersey.

Robert, C.P., C. G. (1999). *Monte Carlo Statistical Methods*. Springer, Berlin Heidelberg New York.

Sanjeev, A., S. Maskell, and N. Gordon (2002). A tutorial on particle filters for online Nonlinear/Non–Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing 50*, 174–188.

Scharth, M. and M. C. Medeiros (2006). Asymmetric effects and long memory in the volatility of Dow Jones stocks. Department of Economics, Pontifical Catholic University of Rio de Janeiro.

Smith, R. L. (1985). Maximum likelihood estimation in a class of non–regular cases. *Biometrika 72*, 67–90.

Smith, R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground–level zone. *Statistical Science 4*, 367–393.

Smith, R. L. and I. Weissman (1994a). Estimating the extremal index. *Journal of the Royal Statistical Society B 56*, 515–528.

Smith, R. L. and I. Weissman (1994b). Estimating the extremal index. *Journal of the Royal Statistical Society B 56*, 515–528.

Tancredi, A., C. Anderson, and A. O'Hagan (2006). Accounting for threshold uncertainty in extreme value estimation. *Extremes 9*, 87–106.

Tawn, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika 77*, 245–253.

Victoria-Feser, M. and E. M. Ronchetti (1994). Robust methods for personal income models. *Canadian Journal of Statistics 22*, 247–258.

Weissman, I. and U. Cohen (1995). The extremal index and clustering of high values for derived stationary sequences. *Journal of Applied Probability 32*, 972–981.

Weissman, I. and S. Y. Novak (1998). On blocks and runs estimators of the extremal index. *Journal of Statistical Planning and Inference 66*, 281–288.

Zhao, X., Q. Hou, D. Lee, and et al. (2007). A comparison between alternative volatility estimations-application on blood oxygen concentration of preterm infants. Modelling and Simulation Society of Australia and New Zealand: MODSIM 2007 International Congress on Modelling and Simulation.

Zhao, X., L. Oxley, C. J. Scarrott, and M. Reale (2009). Extreme value GARCH modelling with Bayesian inference. In *Proc. of MODSIM2009 conference*. Also Department of Economics, University of Canterbury, Working Paper: 05/2009. http://www.econ.canterbury.ac.nz/RePEc/cbt/econwp/0905.pdf.

Zhao, X., C. J. Scarrott, L. Oxley, and M. Reale (2010a). Extreme value modelling for forecasting market crisis impacts. *Apllied Financial Economics 20*, 63–72.

Zhao, X., C. J. Scarrott, L. Oxley, and M. Reale (2010b). GARCH dependence in extreme value model with Bayesian inference. *To appear in Mathematics and Computers in Simulation*.

Zhao, X., C. J. Scarrott, M. Reale, and L. Oxley (2009a). Bayesian extreme value mixture modelling for estimating Value-at-Risk. Technical report, Department of Economics, University of Canterbury, Working Paper.

Zhao, X., C. J. Scarrott, M. Reale, and L. Oxley (2009b). Bayesian extreme value mixture modelling for estimating VaR. *preprint*.